ASESORES TECNICOS

Carlo Federici Casa
Doctor en Física y Matemáticas
Universidad de Génova, Italia
Profesor de Matemáticas
Universidad Nacional, Bogotá

Enzo R. Gentile
Doctor en Matemáticas
Universidad Nacional de Cuyo, Argentina
Estudios de Postgraduado en la Universidad de Princeton, EE. UU.
Ex profesor de Algebra en las Universidades de Rutgers y Northwestern, EE. UU.
Profesor titular Facultad de Ciencias Exactas
Universidad de Buenos Aires

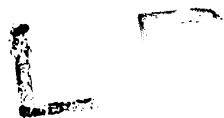
Raúl Bravo F.
Profesor de Matemáticas
Universidad de Chile Santiago

Jesús María Castaño
Profesor de Matemáticas
Universidad del Valle
Cali, Colombia

ASESOR LINGÜISTICO

Dr. Carlos Vega
Profesor de Filología española

Algebra Lineal



KENNETH HOFFMAN
Professor of Mathematics
Massachusetts Institute of Technology

RAY KUNZE
Professor of Mathematics
University of California, Irvine

TRADUCCION Y ADAPTACION

HUGO E. FINSTERBUSCH
Escuela de Graduados, Courant Institute
of Mathematical Science
Master of Science en Matemáticas de la Universidad de N.Y.
Químico. Instituto Politécnico de la Universidad
Católica de Chile
Profesor Asociado, Departamento de Matemáticas Puras
y Aplicadas, Instituto de Matemáticas,
Universidad Católica de Chile

PRENTICE-HALL HISPANOAMERICANA, S. A.

México ▪ Englewood Cliffs ▪ Londres ▪ Sydney ▪ Toronto ▪
Nueva Delhi ▪ Tokio ▪ Singapur ▪ Rio de Janeiro

ALGEBRA LINEAL

PRENTICE-HALL HISPANOAMERICANA, S.A.
Av. San Andrés Atoto 157, Fracc. Industrial San Andrés Atoto
53500 Naucalpan de Juárez, Edo. de México
Miembro de la Cámara Nacional de la Industria Editorial, Reg. Núm. 1524

ISBN 968-880-009-0

Traducido de la segunda edición en inglés de
STATISTICS FOR MANAGEMENT

ISBN 0-13-022046-9

7890123456 I.P.-84 8712345690

Impreso en México Printed In Mexico



MAR

ESTA OBRA SE TERMINO DE IMPRIMIR EN
IMPRESORA AZTECA, S.A.
PONIENTE 140 No. 681-1
MEXICO, D.F.

1000 1991




PREAMBULO

Mucho nos complace que la segunda edición de nuestro libro haya sido traducida al español por el profesor Hugo Finsterbusch. Esperamos que esto pueda contribuir de alguna manera a la formación matemática de los estudiantes del mundo de habla española.

KENNETH HOFFMAN

Prólogo

Nuestro propósito original al escribir este libro fue proporcionar un texto para el curso universitario de álgebra lineal en el Instituto Tecnológico de Massachusetts. Este curso fue diseñado para la especialidad de matemática en el nivel medio, aun cuando tres cuartas partes de los estudiantes eran de otras disciplinas científicas y tecnológicas que iban desde el primer año universitario hasta estudiantes avanzados. Esta descripción del auditorio del MIT para el texto se conserva, en general, hasta el presente exacta. Los diez años transcurridos desde la primera edición han visto la proliferación de cursos de álgebra lineal a través del país, y han brindado a uno de los autores la oportunidad de enseñar la materia básica a una variedad de grupos en la Universidad de Brandeis, Universidad de Washington (St. Louis) y en la Universidad de California (Irvine).

Nuestro principal propósito al revisar el *Algebra Lineal* ha sido incrementar la variedad de cursos que pueden ser dictados con ella. Por un lado hemos estructurado los capítulos, especialmente los más difíciles, de modo que haya varios puntos terminales a lo largo del desarrollo, permitiendo al profesor en un curso de un cuatrimestre o de un semestre ejercitar una cantidad considerable de posibilidades en la elección del tema. Por otra parte hemos aumentado la cantidad de material en el texto, de modo que pueda usarse para un curso anual intensivo en álgebra lineal e incluso como libro de referencia para matemáticos.

Los mayores cambios se han hecho en el tratamiento de las formas canónicas y de los espacios con producto interno. En el capítulo 6 ya no se comienza con la teoría general que fundamenta la teoría de las formas canónicas. Primero se tratan los valores propios en relación con los teoremas de triangulación y diagonalización, y luego se prepara el camino hacia la teoría general. El capítulo 8 se ha dividido de modo que el material básico de espacios con producto interno y diagonalización unitaria sea seguido por el capítulo 9 que trata de las formas sesquilineales y de las propiedades más complicadas de los operadores normales, incluyendo operadores normales en espacios reales con producto interno.

Se ha hecho también un número de pequeños cambios y perfeccionamientos de la primera edición. Pero la doctrina básica que la inspira no ha cambiado.

No hemos concedido atención particular al hecho de que la mayoría de los estudiantes no estén primariamente interesados en matemática, pues creemos que un curso de esta disciplina no debe atiborrar de técnicas al estudiante de ciencia, ingeniería o ciencias sociales, sino procurarle la comprensión de los conceptos matemáticos básicos.

Por otro lado, somos conscientes del amplio campo de conocimientos previos que los estudiantes debieran tener y, en particular, del hecho de que los estudiantes han tenido muy poca experiencia con el razonamiento matemático abstracto. Por esta razón se ha evitado introducir demasiadas ideas abstractas muy al comienzo del libro. Además se ha añadido un Apéndice que introduce o analiza ideas básicas tales como las de conjunto, funciones y relación de equivalencia. Hallamos de mayor utilidad tratar de estas ideas en forma independiente, pero aconsejando a los estudiantes leer el Apéndice cuando ellas aparezcan.

A lo largo del libro se ha incluido una gran variedad de ejemplos de los conceptos importantes que aparecen. El estudio de todos los ejemplos es de fundamental importancia y tiende a minimizar el número de estudiantes que repiten definiciones, teoremas y demostraciones en orden lógico sin comprender el significado de los conceptos abstractos. Este libro contiene también una amplia variedad de ejercicios graduados (alrededor de seiscientos) que varían desde aplicaciones rutinarias hasta otros dirigidos a los mejores estudiantes. Estos ejercicios pretenden ser una parte importante del texto.

El capítulo 1 se refiere a los sistemas de ecuaciones y sus soluciones mediante operaciones elementales por filas de las matrices. Ha sido nuestra práctica dedicar alrededor de seis clases a esta materia. Ello muestra a los estudiantes un bosquejo de los orígenes del álgebra lineal y una técnica de cálculo necesaria para comprender los ejemplos de las ideas abstractas que aparecen en los capítulos posteriores. El capítulo 2 se refiere a los espacios vectoriales, subespacios, bases y dimensión. El capítulo 3 trata las transformaciones lineales, su álgebra y su representación por medio de matrices, como también, del isomorfismo, de funciones lineales y de espacios duales. El capítulo 4 define el álgebra de los polinomios sobre un cuerpo, los ideales en tal álgebra y la factorización prima de un polinomio. También trata sobre raíces, fórmula de Taylor y fórmula de interpolación de Lagrange. El capítulo 5 desarrolla los determinantes de matrices cuadradas y el determinante visto como una función alternada n-lineal de las filas de una matriz, para seguir con las funciones multilineales en módulos como el anillo de Grassman. Lo referente a módulos coloca el concepto de determinante en un marco más amplio y completo que el que usualmente se encuentra en textos elementales. Los capítulos 6 y 7 contienen un estudio de los conceptos que son básicos para el análisis de una transformación lineal en un espacio vectorial de dimensión finita, el análisis de valores propios, las transformaciones triangulables y diagonalizables, los conceptos de las partes diagonalizables y nilpotentes de una transformación general y las formas canónicas racional y de Jordan. Los teoremas de descomposición primaria y cíclica juegan un papel central; a la última se llega a través del estudio de los subespacios admisibles. El capítulo 7 incluye un análisis de las matrices sobre un dominio polinomial, el cálculo de factores invariantes y divisores elementales de una matriz, y el desarrollo de la forma canónica de Smith. El capítulo termina con un estudio sobre los operadores semisimples, para completar el análisis del caso de un operador. El capítulo 8 trata con algún detalle los espacios con producto interno de dimensión finita y cubre la geometría básica, relacionando la ortogonalización con la idea de la «mejor aproximación a un vector» y con los conceptos de proyección ortogonal de un vector sobre un subespacio y el complemento ortogonal de un subespacio. También trata de los operadores unitarios, y culmina con la diagonalización de operadores autoadjuntos y normales. El capítulo 9 introduce las formas sesquilineales, las relaciona con los operadores positivos y autoadjuntos en un espacio con producto interno, sigue con la teoría espectral de operadores normales y luego en resultados más complejos relativos a operadores normales en espacios reales o complejos con producto interno. El capítulo 10 examina las formas bilineales, insistiendo en las formas canónicas para las formas simétrica y antisimétrica, así como en los grupos que preservan formas no-degeneradas, especialmente los grupos ortogonal, unitario, seudo-ortogonal y de Lorentz

Se estima que cualquier curso que use este texto deberá cubrir completamente los capítulos 1, 2 y 3 con posible excepción de las secciones 3.6 y 3.7 que se refieren al doble dual y a la traspuesta de una trasformación lineal. Los capítulos 4 y 5, sobre polinomios y determinantes, pueden tratarse con grados diversos de minuciosidad. En efecto, los ideales de polinomios y las propiedades básicas de los determinantes pueden cubrirse en forma bastante esquemática sin afectar el desarrollo lógico del texto; sin embargo, nos inclinamos a tratar estos capítulos cuidadosamente (excepto los resultados referente a módulos), porque la materia ilustra muy bien las ideas básicas del álgebra lineal. Un curso elemental puede ser ahora concluido elegantemente con las cuatro primeras secciones del capítulo 6, junto con el capítulo 8 (nuevo). Si las formas racional y de Jordan han de ser incluidas, es necesario abarcar una extensión mayor del capítulo 6.

Quedamos muy reconocidos a todos aquéllos que contribuyeron a la primera edición, especialmente a los profesores señores Harry Furstenberg, Louis Howard, Daniel Kan y Edward Thorp, a las señoras Judith Bowers, Betty Ann (Sargent) Rose y a la señorita Phyllis Ruby. Queremos además dar las gracias a los numerosos estudiantes y colegas cuyos penetrantes comentarios llevaron a esta revision, y al personal de Prentice-Hall por su paciencia para tratar con dos autores atrapados en los laberintos de la administración académica. Finalmente, gratitud especial debemos a la señora Sophia Koulouras por su pericia y agotador esfuerzo en escribir a máquina el manuscrito revisado.

K. HOFFMAN y R. KUNZE

Tabla de materias

1. Ecuaciones lineales

1.1. Cuerpos

Suponemos al lector familiarizado con el álgebra elemental de los números reales y complejos. En una gran parte de este libro las propiedades algebraicas de los números que se usarán se deducen fácilmente de la siguiente breve lista de propiedades de la adición y de la multiplicación. Se designa por F el conjunto de los números reales o el conjunto de los números complejos.

1. La adición es conmutativa,

$$x + y = y + x$$

para cualquiera x e y de F.

2. La adición es asociativa,

$$x + (y + z) = (x + y) + z$$

para cualquiera x, y y z de F.

3. Existe un elemento único 0 (cero) de F tal que $x + 0 = x$, para todo x en F.

4. A cada x de F corresponde un elemento único $(-x)$ de F tal que $x + (-x) = 0$.

5. La multiplicación es conmutativa,

$$xy = yx$$

para cualquiera x e y de F.

6. La multiplicación es asociativa,

$$x(yz) = (xy)z$$

para cualquiera x, y y z de F.

7. Existe un elemento no nulo único de F tal que $x1 = x$, para todo x de F.
8. A cada elemento no nulo x de F corresponde un único elemento x^{-1} (o $1/x$) de F tal que $xx^{-1} = 1$.
9. La multiplicación es distributiva respecto de la adición; esto es, $x(y + z) = xy + xz$, para cualesquiera x, y y z de F.

Supóngase que se tiene un conjunto F de objetos $x, y, z, \dots$ y dos operaciones sobre los elementos de F como sigue; la primera operación, llamada adición, asocia a cada par de elementos x, y de F un elemento $(x + y)$ de F; la segunda operación, llamada multiplicación, asocia a cada par x, y de F un elemento xy de F; y estas dos operaciones satisfacen las anteriores condiciones (1)-(9). El conjunto F, junto con estas operaciones, se llama entonces **cuerpo**. Hablando aproximadamente, un cuerpo es un conjunto, junto con algunas operaciones sobre los elementos de éste, que se comportan como la adición, sustracción, multiplicación y división corrientes de los números en el sentido de que obedecen a las nueve reglas del álgebra dadas anteriormente. Con las operaciones comunes de adición y multiplicación, el conjunto C de los números complejos es un cuerpo, como lo es el conjunto R de los números reales.

En la mayor parte de este libro los «números» que se usan pueden ser los elementos de cualquier cuerpo F. Para permitir esta generalidad se usará la palabra «escalar» en vez de «número». No perderá mucho el lector si siempre presupone que el cuerpo de los escalares es un subcuerpo del cuerpo de los números complejos. Un **subcuerpo** de un cuerpo C es un conjunto F de números complejos que es a su vez un cuerpo respecto de las operaciones usuales de adición y multiplicación de números complejos. Esto significa que el 0 y el 1 están en el conjunto F, y que si x e y son elementos de F, también lo son $(x + y)$, $-x$, xy, e x^{-1} (si $x \neq 0$). Un ejemplo de un subcuerpo semejante es el cuerpo R de los números reales; en efecto, si se identifican los números reales con los números complejos $(a + ib)$ para los que $b = 0$, el 0 y el 1 del cuerpo complejo son números reales y si x e y son reales, también lo son $(x + y)$, $-x$, xy y x^{-1} (si $x \neq 0$). Daremos otros ejemplos más adelante. Lo peculiar de los subcuerpos en nuestro estudio es esencialmente lo siguiente: Si se está operando con escalares que forman un cierto subcuerpo de C, entonces la ejecución de las operaciones de adición, sustracción, multiplicación o división con estos escalares no se salen del subcuerpo dado.

Ejemplo 1. El conjunto de los **enteros positivos***: $1, 2, 3, \dots$ no es un subcuerpo de C por varias razones. Por ejemplo, 0 no es un entero positivo; para ningún entero positivo n, es $-n$ un entero positivo; para ningún entero positivo n, excepto 1, es $1/n$ un entero positivo.

Ejemplo 2. El conjunto de los **enteros**: $\dots, -2, -1, 0, 1, 2, \dots$, no es un subcuerpo de C, porque para un entero n, $1/n$ no es entero al menos que

* El autor llama «enteros positivos» los enteros 1, 2, 3, ... y excluye el 0. Hoy no es esto asi, pues se incluye el 0 entre los enteros positivos («naturales»).

n sea 1 o -1. Con las operaciones usuales de adición y multiplicación, el conjunto de los enteros satisface todas las condiciones (1)-(9), con excepción de la condición (8).

Ejemplo 3. El conjunto de los **números racionales,** esto es, números de la forma p/q, donde p y q son enteros y $q \neq 0$, es un subcuerpo del cuerpo de los complejos. La división que no es posible en el conjunto de los enteros es posible en el conjunto de los números racionales. El lector interesado debería verificar que cualquier subcuerpo de C debe contener a todo número racional.

Ejemplo 4. El conjunto de todos los números complejos de la forma $x+y\sqrt{2}$, donde x e y son racionales, es un subcuerpo de C. Se deja al lector la comprobación.

En los ejemplos y ejercicios de este libro el lector deberá suponer que el cuerpo considerado es un subcuerpo de los números complejos, a menos que expresamente se establezca que es un cuerpo más general. No queremos volver sobre este punto; sin embargo, se debe indicar por qué se adopta tal supuesto. Si F es un cuerpo, es posible a veces sumar la unidad 1 a sí misma un número finito de veces y obtener 0 (véase el Ejercicio 5 de la siguiente Sección 1.2):

$$1 + 1 + \cdots + 1 = 0.$$

Esto no sucede en el cuerpo de los números complejos (ni en ningún subcuerpo suyo). Si ello sucede en F, entonces el menor n, tal que la suma de los n unos es 0, se llama **característica** del cuerpo F. Si ello no sucede en F, entonces (por alguna extraña razón) F se llama un cuerpo de **característica cero.** A menudo, cuando se supone que F es un subcuerpo de C, lo que se quiere garantizar es que F es un cuerpo de característica cero; pero, en una primera exposición del álgebra lineal, es preferible no preocuparse mucho con respecto a características de cuerpos.

1.2. Sistemas de ecuaciones lineales

Supóngase que F es un cuerpo. Se considera el problema de encontrar n escalares (elementos de F) $x_1, \ldots, x_n$ que satisfagan las condiciones

$$\begin{array}{ccccccccc} A_{11}x_1 & + & A_{12}x_2 & + & \cdots & + & A_{1n}x_n & = & y_1 \\ A_{21}x_1 & + & A_{22}x_2 & + & \cdots & + & A_{2n}x_n & = & y_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ A_{m1}x_1 & + & A_{m2}x_2 & + & \cdots & + & A_{mn}x_n & = & y_m \end{array} \tag{1-1}$$

donde $y_1, \ldots, y_m$ y A_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, son elementos de F. A (1-1) se le llama un **sistema de m ecuaciones lineales con n incógnitas.** Todo n-tuple $(x_1, \ldots, x_n)$ de elementos de F que satisface cada una de las ecuaciones de (1-1) se llama una **solución** del sistema. Si $y_1 = y_2 = \cdots = y_m = 0$, se dice que el sistema es **homogéneo,** o que cada una de las ecuaciones es homogénea.

Tal vez, la técnica fundamental para encontrar las soluciones de un sistema de ecuaciones lineales, es la técnica de eliminación. Se puede ilustrar esta técnica en el sistema homogéneo

$$\begin{aligned} 2x_1 - x_2 + x_3 &= 0 \\ x_1 + 3x_2 + 4x_3 &= 0. \end{aligned}$$

Si sumamos (-2) veces la segunda ecuación a la primera, se obtiene

$$-7x_2 - 7x_3 = 0$$

o sea, $x_2 = -x_3$. Si se suma tres veces la primera ecuación a la segunda, se obtiene

$$7x_1 + 7x_3 = 0$$

o sea, $x_1 = -x_3$. Así, se concluye que si (x_1, x_2, x_3) es una solución entonces $x_1 = x_2 = -x_3$. A la inversa, se puede fácilmente verificar que toda terna de esa forma es una solución. Así, el conjunto de soluciones consta de todas las ternas $(-a, -a, a)$.

Se han encontrado las soluciones de este sistema de ecuaciones «eliminando incógnitas», esto es, multiplicando las ecuaciones por escalares y sumándolas luego para producir ecuaciones en las cuales algunas de las x_j no están presentes. Deseamos formalizar ligeramente este proceso de modo que se pueda entender por qué opera y para así poder llevar a cabo los cálculos necesarios para resolver un sistema de manera sistemática.

Para el sistema general (1-1), supóngase que seleccionamos m escalares $c_1, \ldots, c_m$, que se multiplica la j-ésima ecuación por c_j y que luego se suma. Se obtiene la ecuación

$$(c_1A_{11} + \cdots + c_mA_{m1})x_1 + \cdots + (c_1A_{1n} + \cdots + c_mA_{mn})x_n = c_1y_1 + \cdots + c_my_m.$$

A tal ecuación se la llama **combinación lineal** de las ecuaciones (1-1). Evidentenente, cualquier solución de todo el sistema de ecuaciones (1-1) será también solución de esta nueva ecuación. Esta es la idea fundamental del proceso de eliminación. Si se tiene otro sistema de ecuaciones lineales

$$\text{(1-2)} \qquad \begin{aligned} B_{11}x_1 + \cdots + B_{1n}x_n &= z_1 \\ \vdots \qquad\quad & \vdots \quad \vdots \\ B_{k1}x_1 + \cdots + B_{kn}x_n &= z_k \end{aligned}$$

en que cada una de las k ecuaciones sea combinación lineal de las ecuaciones de (1-1), entonces toda solución de (1-1) es solución de este nuevo sistema. Evidentemente puede suceder que algunas soluciones de (1-2) no sean soluciones de (1-1). Esto, claro está, no sucede si cada ecuación en el sistema original es combinación lineal de las ecuaciones en el nuevo sistema. Se dirá que dos sistemas de ecuaciones lineales son **equivalentes** si cada ecuación de cada sistema es combinación lineal de las ecuaciones del otro sistema. Podemos entonces establecer formalmente estas observaciones como sigue

Teorema 1. *Sistemas equivalentes de ecuaciones lineales tienen exactamente las mismas soluciones.*

Si se quiere que el proceso de eliminación para encontrar las soluciones de un sistema como (1-1) sea efectivo, entonces se debe buscar, formando combinaciones lineales de las ecuaciones dadas, cómo se obtiene un sistema equivalente de ecuaciones que sea más fácil de resolver. En la próxima sección se estudiará un método para hacerlo.

Ejercicios

1. Verificar que el conjunto de números complejos descritos en el Ejemplo 4 es un subcuerpo de *C*.

2. Sea *F* el cuerpo de los números complejos. ¿Son equivalentes los dos sistemas de ecuaciones lineales siguientes? Si es así, expresar cada ecuación de cada sistema como combinación lineal de las ecuaciones del otro sistema.

$$\begin{aligned} x_1 - x_2 &= 0 \qquad & 3x_1 + x_2 &= 0 \\ 2x_1 + x_2 &= 0 \qquad & x_1 + x_2 &= 0 \end{aligned}$$

3. Examine los siguientes sistemas de ecuaciones como en el Ejercicio 2.

$$\begin{aligned} -x_1 + x_2 + 4x_3 &= 0 \qquad & x_1 \qquad - x_3 &= 0 \\ x_1 + 3x_2 + 8x_3 &= 0 \qquad & x_2 + 3x_3 &= 0 \\ \tfrac{1}{2}x_1 + x_2 + \tfrac{5}{2}x_3 &= 0 \end{aligned}$$

4. Examine los siguientes sistemas como en el Ejercicio 2.

$$\begin{aligned} 2x_1 + (-1+i)x_2 \qquad + x_4 &= 0 \qquad & \left(1+\frac{i}{2}\right)x_1 + 8x_2 - ix_3 - x_4 &= 0 \\ 3x_2 - 2ix_3 + 5x_4 &= 0 \qquad & \tfrac{2}{3}x_1 - \tfrac{1}{2}x_2 + x_3 + 7x_4 &= 0 \end{aligned}$$

5. Sea *F* un conjunto que contiene exactamente dos elementos, 0 y 1. Se define una adición y multiplicación por las tablas:

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

Verificar que el conjunto *F*, juntamente con estas dos operaciones, es un cuerpo.

6. Demostrar que si dos sistemas homogéneos de ecuaciones lineales con dos incógnitas tienen las mismas soluciones, son equivalentes.

7. Demostrar que todo subcuerpo del cuerpo de los números complejos contiene a todo número racional.

8. Demostrar que todo cuerpo de característica cero contiene una copia del cuerpo de los números racionales

1.3. Matrices y operaciones elementales de fila

No se puede menos de advertir que en la formación de combinaciones lineales de ecuaciones lineales no hay necesidad de seguir escribiendo las «incógnitas» $x_1, \ldots, x_n$, ya que realmente solo se opera con los coeficientes A_{ij} y los escalares y_i. El sistema (1-1) se abreviará ahora así:

$$AX = Y$$

donde

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}$$

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{y} \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

A se llama **matriz de los coeficientes** del sistema. Estrictamente hablando, la disposición rectangular expuesta no es una matriz, sino una representación de una matriz. Una **matriz $m \times n$ sobre el cuerpo F** es una función A del conjunto de los pares de enteros (i, j), $1 \leq i \leq m$, $1 \leq j \leq n$, en el cuerpo F. Los **elementos** de la matriz A son los escalares $A(i, j) = A_{ij}$, y, con frecuencia, suele ser más conveniente describir la matriz disponiendo sus elementos en un arreglo rectangular con m filas y n columnas, como antes. Así, X (anteriormente) es, o define, una matriz $n \times 1$, e Y una matriz $m \times 1$. Por el momento, $AX = Y$ no es más que una notación abreviada del sistema de ecuaciones lineales. Más adelante, cuando se haya definido una multiplicación de matrices, querrá decir que Y es el producto de A y X.

Deseamos ahora considerar operaciones sobre las filas de la matriz A que correspondan a la formación de combinaciones lineales de las ecuaciones del sistema $AX = Y$. Se limitará nuestra atención a tres **operaciones elementales de filas** en una matriz $m \times n$, sobre el cuerpo F:

1. Multiplicación de una fila de A por un escalar c no nulo;
2. Remplazo de la r-ésima fila de A por la fila r más c veces la fila s, donde c es cualquier escalar y $r \neq s$;
3. Intercambio de dos filas de A.

Una operación elemental de filas es, pues, un tipo especial de función (regla) e que asocia a cada matriz $m \times n$, A, una matriz $m \times n$, $e(A)$. Se puede describir e en forma precisa en los tres casos como sigue:

1. $e(A)_{ij} = A_{ij}$ si $i \neq r$, $e(A)_{rj} = cA_{rj}$.
2. $e(A)_{ij} = A_{ij}$ si $i \neq r$, $e(A)_{rj} = A_{rj} + cA_{sj}$.
3. $e(A)_{ij} = A_{ij}$ si i es diferente de r y s, $e(A)_{rj} = A_{sj}$, $e(A)_{sj} = A_{rj}$.

En la definición de $e(A)$ no es realmente importante cuántas columnas tenga A, pero el número de filas de A sí que es decisivo. Por ejemplo, se debe prestar

atención a lo que significa intercambiar las filas 5 y 6 de una matriz 5×5. Para evitar esta clase de complicaciones, convendremos en que una operación elemental de filas e está definida en la clase de todas las matrices $m \times n$ sobre F, para algún m fijo, pero para cualquier n. En otras palabras, una e particular está definida en la clase de todas las matrices sobre F que tienen m filas.

Una razón por la que nos limitamos a solo estos tres tipos de operaciones con filas, es que efectuada una operación e en una matriz A, se puede volver a A efectuando una operación semejante en $e(A)$.

Teorema 2. *A cada operación elemental de filas e corresponde una operación elemental de filas* e_1, *del mismo tipo de e, tal que* $e_1(e(A)) = e(e_1(A)) = A$ *para todo A. Es decir, existe la operación (función) inversa de una operación elemental de filas y es una operación elemental de filas del mismo tipo.*

Demostración. (1) Supóngase que e es la operación que multiplica la r-ésima fila de una matriz por un escalar no nulo c. Sea e_1 la operación que multiplica la fila r por c^{-1}. (2) Supóngase que e sea la operación que remplaza la fila r por la misma fila r a la que se le sumó la fila s multiplicada por c, $r \neq s$. Sea e_1 la operación que remplaza la fila r por la fila r a la que se le ha sumado la fila s multiplicada por $(-c)$. (3) Si e intercambia las filas r y s, sea $e_1 = e$. En cada uno de estos casos es claro que $e_1(e(A)) = e(e_1(A)) = A$ para todo A. ▌

Definición. *Si A y B son dos matrices* $m \times n$ *sobre el cuerpo F, se dice que B es* **equivalente por filas** *a A si B se obtiene de A por una sucesión finita de operaciones elementales de filas.*

Usando el Teorema 2, el lector encontrará fácil verificar lo siguiente: Cada matriz es equivalente por filas a ella misma; si B es equivalente por filas a A, entonces A es equivalente por filas a B; si B es equivalente por filas a A y C es equivalente por filas a B, entonces C es equivalente por filas a A. O sea, que la equivalencia por filas es una relación de equivalencia (véase Apéndice).

Teorema 3. *Si A y B son matrices equivalentes por filas, los sistemas homogéneos de ecuaciones lineales* $AX = 0$ *y* $BX = 0$ *tienen exactamente las mismas soluciones.*

Demostración. Supóngase que se pasa de A a B por una sucesión finita de operaciones elementales de filas:

$$A = A_0 \rightarrow A_1 \rightarrow \cdots \rightarrow A_k = B.$$

Basta demostrar que los sistemas $A_jX = 0$ y $A_{j+1}X = 0$ tienen las mismas soluciones, es decir, que una operación elemental por filas no altera el conjunto de soluciones.

Así, supóngase que B se obtiene de A por una sola operación elemental de filas. Sin que importe cuál de los tres tipos (1), (2) o (3) de operaciones sea,

cada ecuación del sistema $BX = 0$ será combinación lineal de las ecuaciones del sistema $AX = 0$. Dado que la inversa de una operacion elemental de filas es una operación elemental de filas, toda ecuación de $AX = 0$ será también combinación lineal de las ecuaciones de $BX = 0$. Luego estos dos sistemas son equivalentes y, por el Teorema 1, tienen las mismas soluciones. ∎

Ejemplo 5. Supóngase que F es el cuerpo de los números racionales

$$A = \begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix}.$$

Se efectuará una sucesión finita de operaciones elementales de filas en A, indicando con números entre paréntesis el tipo de operación realizada.

$$\begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & -2 & -1 & 7 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & \frac{15}{2} & -\frac{55}{2} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(1)}$$

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & -\frac{5}{3} \end{bmatrix}$$

La equivalencia por filas de A con la última matriz en la sucesión anterior nos indica en particular que las soluciones de

$$\begin{aligned} 2x_1 - x_2 + 3x_3 + 2x_4 &= 0 \\ x_1 + 4x_2 \qquad - x_4 &= 0 \\ 2x_1 + 6x_2 - x_3 + 5x_4 &= 0 \end{aligned}$$

y

$$\begin{aligned} x_3 - \tfrac{11}{3}x_4 &= 0 \\ x_1 \qquad + \tfrac{17}{3}x_4 &= 0 \\ x_2 \qquad - \tfrac{5}{3}x_4 &= 0 \end{aligned}$$

son exactamente las mismas. Queda de manifiesto en el segundo sistema que, si asignamos a x_4 un valor racional c cualquiera, se obtiene una solución $(-\frac{17}{3}c, \frac{5}{3}c, \frac{11}{3}c, c)$, con lo que toda solución es de esta forma

Ejemplo 6. Supóngase que es F el cuerpo de los números complejos y

$$A = \begin{bmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{bmatrix}.$$

Al efectuar operaciones con filas es conveniente a menudo combinar varias operaciones del tipo (2). Teniendo esto presente

$$\begin{bmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 2+i \\ 0 & 3+2i \\ 1 & 2 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & 1 \\ 0 & 3+2i \\ 1 & 2 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Con lo que el sistema de ecuaciones

$$\begin{aligned} -x_1 + ix_2 &= 0 \\ -ix_1 + 3x_2 &= 0 \\ x_1 + 2x_2 &= 0 \end{aligned}$$

tiene solamente la solución trivial $x_1 = x_2 = 0$.

En los Ejemplos 5 y 6 obviamente no se estaban haciendo operaciones con filas al azar. La elección de las operaciones con filas estaba motivada por el deseo de simplificar la matriz coeficiente de una manera análoga a la «eliminación de incógnitas» en el sistema de ecuaciones lineales. Daremos ahora una definición formal del tipo de matriz a la que estábamos tratando de llegar.

Definición. *Una matriz $m \times n$, R, se llama* **reducida por filas** *si:*

(a) *el primer elemento no nulo de cada fila no nula de R es igual a* 1;

(b) *cada columna de R que tiene el primer elemento no nulo de alguna fila tiene todos sus otros elementos* 0.

Ejemplo 7. Un ejemplo de matriz reducida por filas es la **matriz identidad** $n \times n$ (cuadrada) I. Esta es la matriz $n \times n$ definida por

$$I_{ij} = \delta_{ij} = \begin{cases} 1, & \text{si} \quad i = j \\ 0, & \text{si} \quad i \neq j. \end{cases}$$

Esta es la primera de las muchas ocasiones en que se usará la **delta de Kronecker** (δ).

En los Ejemplos 5 y 6 las matrices finales obtenidas son matrices reducidas por filas. Dos ejemplos de matrices *no* reducidas por filas son:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix}.$$

La segunda matriz no satisface la condición (a), porque el primer elemento

no nulo del primer renglón no es 1. La primera matriz sí satisface la condición (a), pero no la condición (b) en la columna 3.

Demostraremos ahora que se puede pasar de cualquier matriz dada a una matriz reducida por filas, por medio de un número finito de operaciones elementales de fila. Esto, junto con el Teorema 3, proporcionará un instrumento efectivo para la resolución de sistemas de ecuaciones lineales.

Teorema 4. *Toda matriz $m \times n$ sobre el cuerpo F es equivalente por filas a una matriz reducida por filas.*

Demostración. Sea A una matriz $m \times n$ sobre F. Si todo elemento de la primera fila de A es 0, la condición (a) se cumple en lo que concierne a la fila 1. Si la fila 1 tiene un elemento no nulo, sea k el menor entero positivo j para el que $A_{1j} \neq 0$. Multiplicando la fila 1 por A_{1k}^{-1}, la condición (a) se cumple con respecto a esa fila. Luego, para todo $i \geq 2$, se suma $(-A_{ik})$ veces la fila 1 a la fila i. Y ahora el primer elemento no nulo de la fila 1 está en la columna k, ese elemento es 1, y todo otro elemento de la columna k es 0.

Considérese ahora la matriz que resultó de lo anterior. Si todo elemento de la fila 2 es 0, se deja tal cual. Si algún elemento de la fila 2 es diferente de 0, se multiplica esa fila por un escalar de modo que el primer elemento no nulo sea 1. En el caso de que la fila 1 haya tenido un primer elemento no nulo en la columna k, este primer elemento no nulo de la fila 2 no puede estar en la columna k; supóngase que esté en la columna $k_r \neq k$. Sumando múltiplos apropiados de la fila 2 a las otras filas, se puede lograr que todos los elementos de la columna k_r sean 0, excepto el 1 en la fila 2. Lo que es importante observar es lo siguiente: Al efectuar estas operaciones, no se alteran los elementos de la fila 1 en las columnas $1, \ldots, k$, ni ningún elemento de la columna k. Es claro que, si la fila 1 era idénticamente nula, las operaciones con la fila 2 no afectan la fila 1.

Si se opera, como se indicó, con una fila cada vez, es evidente que después de un número finito de etapas se llegará a una matriz reducida por filas. ∎

Ejercicios

1. Hallar todas las soluciones del sistema de ecuaciones

$$\begin{aligned}(1 - i)x_1 - ix_2 &= 0\\ 2x_1 + (1 - i)x_2 &= 0.\end{aligned}$$

2. Si

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{bmatrix}$$

hallar todas las soluciones de $AX = 0$ reduciendo A por filas.

3. Si

$$A = \begin{bmatrix} 6 & -4 & 0 \\ 4 & -2 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

hallar todas las soluciones de $AX = 2X$ y todas las soluciones de $AX = 3X$ (el símbolo cX representa la matriz , cada elemento de la cual es c veces el correspondiente elemento de X).

4. Hallar una matriz reducida por filas que sea equivalente por filas a

$$A = \begin{bmatrix} i & -(1+i) & 0 \\ 1 & -2 & 1 \\ 1 & 2i & -1 \end{bmatrix}.$$

5. Demostrar que las siguientes dos matrices *no* son equivalentes por filas

$$\begin{bmatrix} 2 & 0 & 0 \\ a & -1 & 0 \\ b & c & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 2 \\ -2 & 0 & -1 \\ 1 & 3 & 5 \end{bmatrix}.$$

6. Sea

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

una matriz 2×2 con elementos complejos. Supóngase que A es reducida por filas y también que $a + b + c + d = 0$. Demostrar que existen exactamente tres de estas matrices.

7. Demostrar que el intercambio de dos filas en una matriz puede hacerse por medio de un número finito de operaciones elementales con filas de los otros dos tipos.

8. Considerar el sistema de ecuaciones $AX = 0$, donde

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

es una matriz 2×2 sobre el cuerpo F. Demostrar lo siguiente:

(a) Si todo elemento de A es 0, entonces cada par (x_1, x_2) es una solución de $AX = 0$.
(b) Si $ad - bc \neq 0$, el sistema $AX = 0$ tiene solamente la solución trivial $x_1 = x_2 = 0$.
(c) Si $ad - bc = 0$ y algún elemento de A es diferente de 0, entonces existe una solución (x_1^0, x_2^0) tal que (x_1, x_2) es una solución si, y solo si, existe un escalar y tal que $x_1 = yx_1^0$, $x_2 = yx_2^0$.

1.4. Matrices escalón reducidas por filas

Hasta el momento, el operar con sistemas de ecuaciones estaba motivado por tratar de encontrar las soluciones de tal sistema. En la Sección 1.3 se estableció un método normal de encontrar esas soluciones. Queremos ahora adquirir alguna información algo más teórica, y para tal propósito es conveniente profundizar un poco más en las matrices reducidas por fila.

Definición. *Una matriz $m \times n$, R, se llama* **matriz escalón reducida por fila** *si:*

(a) *R es reducida por filas;*
(b) *toda fila de R que tiene todos sus elementos 0 está debajo de todas las filas que tienen elementos no nulos;*

(c) *si las filas* 1, ..., r *son las filas no nulas de* R, *y si el primer elemento no nulo de la fila i está en la columna* k_i, $i = 1, \ldots, r$, *entoces* $k_1 < k_2 < \cdots < k_r$.

Se puede describir también una matriz escalón R reducida por filas como sigue. Todo elemento de R es 0, o existe un número positivo r, $1 \leq r \leq m$, y r enteros positivos $k_1, \ldots, k_r$ con $1 \leq k_i \leq n$ y

(a) $R_{ij} = 0$ para $i > r$, y $R_{ij} = 0$ si $j < k_i$.
(b) $R_{ik_j} = \delta_{ij}$, $1 \leq i \leq r$, $1 \leq j \leq r$.
(c) $k_1 < \cdots < k_r$.

Ejemplo 8. Dos ejemplos de matrices escalón reducidas por filas son la matriz identidad $n \times n$ y la **matriz cero** $m \times n$, $0^{m,n}$, cuyos elementos son todos 0. El lector no tendrá dificultad en hallar otros ejemplos, pero quisiéramos dar otro no trivial:

$$\begin{bmatrix} 0 & 1 & -3 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Teorema 5. *Toda matriz* $m \times n$, A, *es equivalente por filas a una matriz escalón por filas.*

Demostración. Sabemos que A es equivalente por filas a una matriz reducida por filas. Todo lo que se necesita observar es que, efectuando un número finito de intercambios de filas en una matriz reducida por filas, se la puede llevar a la forma escalón reducida por filas. ▮

En los Ejemplos 5 y 6 se vio la importancia que tienen las matrices reducidas por filas en la resolución de sistemas homogéneos de ecuaciones lineales. Sea ahora examinar brevemente el sistema $RX = 0$, donde R es una matriz escalón reducida por filas. Sean las filas 1, ..., r las no nulas de R, y supóngase que el elemento principal no nulo de la fila i está en la columna k_i. El sistema $RX = 0$ consta entonces de r ecuaciones no triviales. Además, la incógnita x_{k_i} aparecerá (con coeficiente no nulo) solamente en la i-ésima ecuación. Si $u_1, \ldots, u_{n-r}$ representan las $(n - r)$ incógnitas que son diferentes de $x_{k_1}, \ldots, x_{k_r}$, entonces las r ecuaciones no triviales de $RX = 0$ son de la forma

$$\text{(1-3)} \qquad \begin{aligned} x_{k_1} + \sum_{j=1}^{n-r} C_{1j}u_j &= 0 \\ \vdots \qquad & \quad \vdots \\ x_{k_r} + \sum_{j=1}^{n-r} C_{rj}u_j &= 0. \end{aligned}$$

Todas las soluciones del sistema de ecuaciones $RX = 0$ se obtienen dando valores arbitrarios a $u_1, \ldots, u_{n-r}$, y calculando entonces los correspondientes valores de $x_{k_1}, \ldots, x_{k_r}$ de (1-3). Por ejemplo, si R es la matriz del Ejemplo 8.

entonces $r = 2$, $k_1 = 2$, $k_2 = 4$ y las dos ecuaciones no triviales del sistema $RX = 0$ son

$$x_2 - 3x_3 + \tfrac{1}{2}x_5 = 0 \quad \text{o} \quad x_2 = 3x_3 - \tfrac{1}{2}x_5$$
$$x_4 + 2x_5 = 0 \quad \text{o} \quad x_4 = -2x_5.$$

Así, podemos asignar cualquier valor a x_1, x_3 y x_5, digamos $x_1 = a$, $x_3 = b$, $x_5 = c$, y obtener la solución $(a,\ 3b - \frac{1}{2}c,\ b,\ -2c,\ c)$.

Observemos una cosa más, en relación con el sistema de ecuaciones $RX = 0$. Si el número r de filas no nulas de R es menor que n, entonces el sistema $RX = 0$ tiene una solución no trivial, esto es, una solución $(x_1, \ldots, x_n)$ en que no todo x_j es 0. En efecto, como $r < n$, se puede elegir algún x_j que no esté entre las r incógnitas $x_{k_1}, \ldots, x_{k_r}$, y se puede entonces construir una solución como antes, en que este x_j es 1. Esta observación nos lleva a uno de los hechos más fundamentales sobre sistemas homogéneos de ecuaciones lineales.

Teorema 6. *Si A es una matriz $m \times n$ con $m < n$, el sistema homogéneo de ecuaciones lineales $AX = 0$ tiene una solución no trivial.*

Demostración. Sea R una matriz escalón reducida por fila que sea equivalente por fila a A. Entonces los sistemas $AX = 0$ y $RX = 0$ tienen las mismas soluciones por el Teorema 3. Si r es el número de filas no nulas de R, entonces ciertamente $r \leq m$, y como $m < n$ tenemos que $r < n$. Se sigue inmediatamente de las observaciones anteriores que $AX = 0$ tiene una solución no trivial. ∎

Teorema 7. *Si A es una matriz $n \times n$ (cuadrada), A es equivalente por filas a la matriz identidad $n \times n$, si, y solo si, el sistema de ecuaciones $AX = 0$ tiene solamente la solución trivial.*

Demostración. Si A es equivalente por filas a I, entonces $AX = 0$ e $IX = 0$ tienen las mismas soluciones. Recíprocamente, supóngase que $AX = 0$ tiene solamente la solución trivial $X = 0$. Sea R una matriz escalón reducida por filas $n \times n$, que es equivalente por filas a A, y sea r el número de filas no nulas de R. Entonces $RX = 0$ carece de solución no trivial. Con lo que $r \geq n$. Pero como R tiene n filas, ciertamente $r \leq n$. Con lo que $r = n$. Como esto quiere decir que R tiene un 1 como primer elemento no nulo en cada una de sus n filas y como estos 1 están en las diferentes columnas n, R debe ser la matriz identidad $n \times n$. ∎

Preguntémonos ahora qué operaciones elementales de fila hay que hacer para resolver un sistema de ecuaciones lineales $AX = Y$ no homogéneo. De entrada cabe observar una diferencia básica con el caso homogéneo; y es que mientras el sistema homogéneo siempre tiene la solución trivial $x_1 = \cdots = x_n = 0$, un sistema no homogéneo no tiene necesariamente solución.

Se construye la **matriz aumentada** A' del sistema $AX = Y$. Esta es la matriz $m \times (n + 1)$ cuyas primeras n columnas son las columnas de A y cuya última columna es Y; más precisamente,

$$A'_{ij} = A_{ij}, \quad \text{si} \quad j \leq n$$
$$A'_{i(n+1)} = y_i.$$

Supóngase que se efectúa en A una sucesión de operaciones elementales con filas, para llegar a una matriz escalón reducida por filas R. Si se efectúa esta misma sucesión de operaciones de fila en la matriz aumentada A', se llegará a una matriz R' cuyas primeras columnas son las columnas de R y cuya última columna tiene ciertos escalares $z_1, \ldots, z_m$. Los escalares z_i son los elementos de la matriz $m \times 1$

$$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}$$

que resulta de la aplicación de la sucesión de operaciones de fila a la matriz Y. Para el lector es claro que, al igual que en la demostración del Teorema 3, los sistemas $AX = Y$ y $RX = Z$ son equivalentes y, como consecuencia, tienen las mismas soluciones. Es muy fácil determinar cuándo el sistema $RX = Z$ tiene alguna solución y encontrar todas las soluciones si es que existen. En efecto, si R tiene r filas no nulas, con el primer elemento no nulo en la fila i y en la columna k_i, $i = 1, \ldots, r$, entonces las primeras r ecuaciones de $RX = Z$ expresarán efectivamente las $x_{k_1}, \ldots, x_{k_r}$ por las $(n - r)$ restantes x_j y los escalares $z_1, \ldots, z_r$. Las últimas $(m - r)$ ecuaciones son

$$\begin{aligned} 0 &= z_{r+1} \\ \vdots\ & \quad \vdots \\ 0 &= z_m \end{aligned}$$

y, por tanto, la condición para que el sistema tenga una solución es $z_i = 0$ para $i > r$. Si se cumple esta condición todas las soluciones del sistema se encontrarán, del mismo modo que en el caso homogéneo, dando valores arbitrarios a $(n - r)$ de las x_j, calculando luego x_{k_i} en la i-ésima ecuación.

Ejemplo 9. Sea F el cuerpo de los números racionales y

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & 1 & 1 \\ 0 & 5 & -1 \end{bmatrix}$$

Se trata de resolver el sistema $AX = Y$ para ciertos y_1, y_2 e y_3. Efectuando una sucesión de operaciones de fila en la matriz aumentada A', que reduzca por filas a A:

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 2 & 1 & 1 & y_2 \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & 0 & \frac{3}{5} & \frac{1}{5}(y_1 + 2y_2) \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix}$$

La condición para que el sistema $AX = Y$ tenga una solución es, pues,

$$2y_1 - y_2 + y_3 = 0$$

y si los escalares dados y_i satisfacen esta condición, todas las soluciones se obtienen asignando un valor c a x_3, y calculando luego

$$\begin{aligned} x_1 &= -\tfrac{3}{5}c + \tfrac{1}{5}(y_1 + 2y_2) \\ x_2 &= \tfrac{1}{5}c + \tfrac{1}{5}(y_2 - 2y_1). \end{aligned}$$

Una observación final respecto al sistema $AX = Y$. Supóngase que los elementos de la matriz A y los escalares $y_1, \ldots, y_m$ pertenecen a un subcuerpo F_1 del cuerpo F. Si el sistema de ecuaciones $AX = Y$ tiene una solución con $x_1, \ldots, x_n$ en F, tiene también una solución con $x_1, \ldots, x_n$ en F_1. En efecto, sobre cualquier cuerpo, la condición para que el sistema tenga una solución es que se cumplan ciertas relaciones entre los $y_1, \ldots, y_n$ en F_1 (las relaciones $z_i = 0$ para $i > r$ anteriores). Por ejemplo, si $AX = Y$ es un sistema de ecuaciones lineales, en el cual los escalares y_k y A_{ij} son números reales, y si existe una solución en que $x_1, \ldots, x_n$ son números complejos, entonces existe una solución en que los $x_1, \ldots, x_n$ son números reales.

Ejercicios

1. Hallar, mediante reducción por filas de la matriz de coeficientes todas las soluciones del siguiente sistema de ecuaciones:

$$\begin{aligned} \tfrac{1}{3}x_1 + 2x_2 - 6x_3 &= 0 \\ -4x_1 \qquad\quad + 5x_3 &= 0 \\ -3x_1 + 6x_2 - 13x_3 &= 0 \\ -\tfrac{7}{3}x_1 + 2x_2 - \tfrac{8}{3}x_3 &= 0 \end{aligned}$$

2. Hallar una matriz escalón reducida por filas que sea equivalente a

$$A = \begin{bmatrix} 1 & -i \\ 2 & 2 \\ i & 1+i \end{bmatrix}.$$

¿Cuáles son las soluciones de $AX = 0$?

3. Describir explícitamente todas las matrices escalón 2×2 reducidas por filas.

4. Considérese el sistema de ecuaciones

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 1 \\ 2x_1 \qquad + 2x_3 &= 1 \\ x_1 - 3x_2 + 4x_3 &= 2. \end{aligned}$$

¿Tiene este sistema solución? Si es así, determinar todas sus soluciones.

5. Dar un ejemplo de un sistema de dos ecuaciones lineales con dos incógnitas que no tenga solución.

6. Mostrar que el sistema

$$\begin{aligned} x_1 - 2x_2 + x_3 + 2x_4 &= 1 \\ x_1 + x_2 - x_3 + x_4 &= 2 \\ x_1 + 7x_2 - 5x_3 - x_4 &= 3 \end{aligned}$$

no tiene solución.

7. Hallar todas las soluciones de

$$\begin{aligned} 2x_1 - 3x_2 - 7x_3 + 5x_4 + 2x_5 &= -2 \\ x_1 - 2x_2 - 4x_3 + 3x_4 + x_5 &= -2 \\ 2x_1 \qquad - 4x_3 + 2x_4 + x_5 &= 3 \\ x_1 - 5x_2 - 7x_3 + 6x_4 + 2x_5 &= -7. \end{aligned}$$

8. Sea

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{bmatrix}.$$

¿Para cuáles ternas (y_1, y_2, y_3) tiene una solución el sistema $AX = Y$?

9. Sea

$$A = \begin{bmatrix} 3 & -6 & 2 & -1 \\ -2 & 4 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 1 & -2 & 1 & 0 \end{bmatrix}.$$

¿Para cuál (y_1, y_2, y_3, y_4) tiene solución el sistema de ecuaciones $AX = Y$?

10. Supóngase que R y R' son matrices escalón 2×3 reducidas por filas y que los sistemas $RX = 0$ y $R'X = 0$ tienen exactamente las mismas soluciones. Demostrar que $R = R'$.

1.5. Multiplicación de matrices

Es evidente (o debería serlo, en todo caso) que el proceso de formar combinaciones lineales con las filas de una matriz es fundamental. Por esta razón es de provecho introducir un esquema sistemático que indique qué operaciones se han de realizar; más precisamente, supóngase que B es una matriz $n \times p$ sobre el cuerpo F, con filas $\beta_1, \ldots, \beta_n$ y que a partir de B se construye una matriz C con filas $\gamma_1, \ldots, \gamma_m$, efectuando ciertas combinaciones lineales

$$\gamma_i = A_{i1}\beta_1 + A_{i2}\beta_2 + \cdots + A_{in}\beta_n. \tag{1-4}$$

Las filas de C quedan determinadas por los mn escalares A_{ij}, que son los elementos de una matriz $m \times n$, A. Si se desarrolla (1-4)

$$(C_{i1} \cdots C_{ip}) = \sum_{r=1}^{n} (A_{ir}B_{r1} \cdots A_{ir}B_{rp})$$

se ve que los elementos de C vienen dados por

$$C_{ij} = \sum_{r=1}^{n} A_{ir}B_{rj}.$$

Definición. *Sea A una matriz $m \times n$ sobre el cuerpo F y sea B una matriz $n \times p$ sobre F. El* **producto** *AB es la matriz $m \times p$, C, cuyos elementos i, j son*

$$C_{ij} = \sum_{r=1}^{n} A_{ir}B_{rj}.$$

Ejemplo 10. He aquí algunos productos de matrices de elementos racionales.

(a)
$$\begin{bmatrix} 5 & -1 & 2 \\ 0 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 5 & -1 & 2 \\ 15 & 4 & 8 \end{bmatrix}$$

Donde

$$\begin{aligned} \gamma_1 &= (5 \quad -1 \quad 2) = 1 \cdot (5 \quad -1 \quad 2) + 0 \cdot (15 \quad 4 \quad 8) \\ \gamma_2 &= (0 \quad 7 \quad 2) = -3(5 \quad -1 \quad 2) + 1 \cdot (15 \quad 4 \quad 8) \end{aligned}$$

(b)
$$\begin{bmatrix} 0 & 6 & 1 \\ 9 & 12 & -8 \\ 12 & 62 & -3 \\ 3 & 8 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 3 \\ 5 & 4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 1 \\ 3 & 8 & -2 \end{bmatrix}$$

Donde

$$\begin{aligned} \gamma_2 &= (9 \quad 12 \quad -8) = -2(0 \quad 6 \quad 1) + 3(3 \quad 8 \quad -2) \\ \gamma_3 &= (12 \quad 62 \quad -3) = 5(0 \quad 6 \quad 1) + 4(3 \quad 8 \quad -2) \end{aligned}$$

(c)
$$\begin{bmatrix} 8 \\ 29 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

(d)
$$\begin{bmatrix} -2 & -4 \\ 6 & 12 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} [2 \quad 4]$$

Donde

$$\gamma_2 = (6 \quad 12) = 3(2 \quad 4)$$

(e)
$$[2 \quad 4] \begin{bmatrix} -1 \\ 3 \end{bmatrix} = [10]$$

(f)
$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -5 & 2 \\ 2 & 3 & 4 \\ 9 & -1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(g)
$$\begin{bmatrix} 1 & -5 & 2 \\ 2 & 3 & 4 \\ 9 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 9 & 0 \end{bmatrix}$$

Es importante observar que el producto de dos matrices puede no estar definido; el producto está definido si, y solo si, el número de columnas de la primera matriz coincide con el número de filas de la segunda. Así que no tiene sentido intercambiar el orden de los factores en (a), (b) y (c) de los ejemplos anteriores. Frecuentemente se escribirán productos, tales como AB, sin mencionar explícitamente las dimensiones de los factores, y en tales casos se dará por entendido que el producto está definido. Por (d), (e), (f), (g) se ve que, aun-

que los productos AB y BA estén definidos, no es necesariamente $AB = BA$; es decir, la multiplicación de matrices *no* es *conmutativa.*

Ejemplo 11.

(a) Si I es la matriz identidad $m \times m$ y A es una matriz $m \times n$, $IA = A$.
(b) Si I es la matriz identidad $n \times n$ y A es una matriz $m \times n$, $AI = A$.
(c) Si $0^{k,m}$ es la matriz nula $k \times m$, $0^{k,n} = 0^{k,m}A$. En forma similar, $A0^{n,p} = 0^{n,p}$.

Ejemplo 12. Sea A una matriz $m \times n$ sobre F. La representación abreviada anterior $AX = Y$, para sistemas de ecuaciones lineales, es compatible con la definición de productos de matrices. Así, si

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

con x_i en F, entonces AX es la matriz $m \times 1$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

tal que $y_i = A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_n$.

El uso de matrices columna sugiere una representación que es frecuentemente útil. Si B es una matriz $n \times p$, las columnas de B son las matrices $n \times 1$, $B_1, \ldots, B_p$, definidas por

$$B_j = \begin{bmatrix} B_{1j} \\ \vdots \\ B_{nj} \end{bmatrix}, \qquad 1 \leq j \leq p.$$

La matriz B es la sucesión de estas columnas:

$$B = [B_1, \ldots, B_p].$$

El elemento i, j de la matriz producto AB se forma a partir de la i-ésima fila de A y de la j-ésima columna de B. Compruebe el lector que la j-ésima columna de AB es AB_j:

$$AB = [AB_1, \ldots, AB_p].$$

A pesar de que un producto de matrices depende del orden en que los factores están escritos, es independiente del modo en que están asociados, como lo demuestra el siguiente teorema.

Teorema 8. *Si A, B, C son matrices sobre el cuerpo F, tales que los productos BC y $A(BC)$ están definidos, entonces también lo están los productos AB, $(AB)C$ y*

$$A(BC) = (AB)C.$$

Demostración. Supóngase que B es una matriz $n \times p$. Como BC está definida, C es una matriz con p filas, y BC tiene n filas. Como $A(BC)$ está definida, se puede suponer que A es una matriz $m \times n$. Así el producto AB existe y es una matriz $m \times p$, de lo que se sigue que el producto $(AB)C$ existe. Para hacer ver que $A(BC) = (AB)C$ se debe demostrar que

$$[A(BC)]_{ij} = [(AB)C]_{ij}$$

para todos los i, j. Por definición

$$\begin{aligned}
[A(BC)]_{ij} &= \sum_r A_{ir}(BC)_{rj} \\
&= \sum_r A_{ir} \sum_s B_{rs}C_{sj} \\
&= \sum_r \sum_s A_{ir}B_{rs}C_{sj} \\
&= \sum_s \sum_r A_{ir}B_{rs}C_{sj} \\
&= \sum_s (\sum_r A_{ir}B_{rs})C_{sj} \\
&= \sum_s (AB)_{is}C_{sj} \\
&= [(AB)C]_{ij}. \quad \blacksquare
\end{aligned}$$

Si A es una matriz (cuadrada) $n \times n$, el producto AA está definido. Esta matriz se representa por A^2. Por el Teorema 8, $(AA)A = A(AA)$ o $A^2A = AA^2$, de modo que el producto AAA está definido sin ambigüedad. Este producto se representa por A^3. En general, el producto $AA \cdots A$ (k veces) está definido sin ambigüedad, y se representará por A^k.

Obsérvese que la relación $A(BC) = (AB)C$ implica, entre otras cosas, que combinaciones lineales de combinaciones lineales de filas de C son otra vez combinaciones lineales de filas de C.

Si B es una matriz y C se obtiene a partir de B por medio de una operación elemental de filas, entonces toda fila de C es combinación lineal de las filas de B, y, por tanto, existe una matriz A tal que $AB = C$. En general, existen muchas de estas matrices A, y entre todas ellas es conveniente y posible escoger una que tenga algunas propiedades especiales. Antes de ver esto se necesita introducir una clase de matrices.

Definición. *Una matriz $m \times m$ se dice* **matriz elemental** *si se puede obtener de la matriz identidad $m \times m$ por medio de una sola operación elemental simple de filas.*

Ejemplo 13. Una matriz elemental 2×2 es necesariamente una de las siguientes:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$$

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}, \quad c \neq 0, \quad \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}, \quad c \neq 0.$$

Teorema 9. *Sea e una operación elemental de fila y sea E la matriz elemental $m \times m$, $E = e(I)$. Entonces, para toda matriz $m \times n$, A*

$$e(A) = EA.$$

Demostración. La clave de la demostración radica en que el elemento de la i-ésima fila y la j-ésima columna de la matriz producto EA se obtiene de la i-ésima fila de E y de la j-ésima columna de A. Los tres tipos de operaciones elementales de fila deben ser estudiados separadamente. Se dará una demostración detallada para una operación del tipo (ii). Los otros dos casos, más fáciles de estudiar, se dejan como ejercicios. Supóngase que $r \neq s$ y que e es una operación que «remplaza la fila r por la fila r más c veces la fila s'». Entonces

$$E_{ik} = \begin{cases} \delta_{ik}, & i \neq r \\ \delta_{rk} + c\delta_{sk}, & i = r. \end{cases}$$

Luego

$$(EA)_{ij} = \sum_{k=1}^{m} E_{ik}A_{kj} = \begin{cases} A_{ik}, & i \neq r \\ A_{rj} + cA_{sj}, & i = r. \end{cases}$$

Es decir, $EA = e(A)$. ▌

Corolario. *Sean A y B dos matrices $m \times n$ sobre el cuerpo F. Entonces B es equivalente por filas a A si, y solo si, $B = PA$, donde P es un producto de matrices elementales $m \times m$.*

Demostración. Supóngase que $B = PA$, donde $P = E_s \cdots E_2E_1$ y los E_i son matrices elementales $m \times m$. Entonces E_1A es equivalente por filas a A_1 y $E_2(E_1A)$ es equivalente por filas a E_1A. Luego E_2E_1A es equivalente por filas a A, y continuando de este modo se ve que $(E_s \cdots E_1)A$ es equivalente por filas a A. Sean $E_1, E_2, \ldots, E_s$ matrices elementales correspondientes a cierta sucesión de operaciones elementales de filas que lleva A a B. Entonces $B = (E_s \cdots E_1)A$. ▌

Ejercicios

1. Sean

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}, \quad C = [1 \quad -1].$$

Calcular ABC y CAB.

2. Sean

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & 1 \\ 3 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -2 \\ 1 & 3 \\ 4 & 4 \end{bmatrix}.$$

Verificar directamente que $A(AB) = A^2B$.

3. Encontrar dos matrices 2×2, A, diferentes tales que $A^2 = 0$, pero $A \neq 0$.

4. Para cada A del Ejercicio 2, hallar matrices elementales $E_1, E_2, \ldots, E_k$ tales que

$$E_k \cdots E_2 E_1 A = I.$$

5. Sean

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 2 \\ 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 3 & 1 \\ -4 & 4 \end{bmatrix}.$$

¿Existe una matriz C tal que $CA = B$?

6. Sea A una matriz $m \times n$ y B una matriz $n \times k$. Demostrar que las columnas de $C = AB$ son combinaciones lineales de las columnas de A. Si $\alpha_1, \ldots, \alpha_n$ son las columnas de A y $\gamma_1, \ldots, \gamma_k$ son las cdlumnas de C, entonces

$$\gamma_j = \sum_{r=1}^{n} B_{rj}\alpha_r.$$

7. Sean A y B matrices 2×2 tales que $AB = I$. Demostrar que $BA = I$.

8. Sea

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

una matriz 2×2. Se desea saber si es posible encontrar matrices 2×2, A y B, tales que $C = AB - BA$. Demostrar que tales matrices pueden hallarse si, y solo si, $C_{11} + C_{22} = 0$.

1.6. *Matrices inversibles*

Supóngase que P es una matriz $m \times m$, que es un producto de matrices elementales. Para cada matriz $m \times n$, A, la matriz $B = PA$ es equivalente por filas a A; luego A es equivalente por filas a B y existe un producto Q de matrices elementales, tal que $A = QB$. En particular, esto es verdad cuando A es la matriz identidad $m \times m$. En otras palabras, existe una matriz $m \times m$, Q, que es ella misma un producto de matrices elementales, tal que $QP = I$. Como veremos, la existencia de una Q con $QP = I$ es equivalente al hecho de que P es un producto de matrices elementales.

Definición. *Sea A una matriz (cuadrada) $n \times n$ sobre el cuerpo F. Una matriz $n \times n$, B, tal que $BA = I$ se llama* **inversa a la izquierda** *de A; una matriz $n \times n$, B, tal que $AB = I$ se llama* **inversa a la derecha** *de A. Si $AB = BA = I$, entonces B se llama* **inversa bilátera** *de A, y se dice que A es* **inversible.**

Lema. *Si A tiene una inversa a la izquierda, B, y una inversa a la derecha, C, entonces $B = C$.*

Demostración. Supóngase que $BA = I$ y que $AC = I$. Entonces

$$B = BI = B(AC) = (BA)C = IC = C. \quad \blacksquare$$

Así, si A es inversa a la izquierda e inversa a la derecha, A es inversible y tiene una inversa bilátera que se representará por A^{-1} y se llamará simplemente **la inversa** de A.

Teorema 10. *Sean A y B dos matrices $n \times n$ sobre F.*

(i) *Si A es inversible, también lo es A^{-1} y $(A^{-1})^{-1} = A$.*
(ii) *Si A y B son inversibles, también lo es AB y $(AB)^{-1} = B^{-1}A^{-1}$.*

Demostración. La primera afirmación es evidente por la simetría de la definición. La segunda se desprende de las relaciones

$$(AB)(B^{-1}A^{-1}) = (B^{-1}A^{-1})(AB) = I. \quad \blacksquare$$

Corolario. *Un producto de matrices inversibles es inversible.*

Teorema 11. *Una matriz elemental es inversible.*

Demostración. Sea E una matriz fundamental correspondiente a la operación elemental de fila e. Si e_1 es la operación inversa de e (Teorema 2) y $E_1 = e_1(I)$, entonces

$$EE_1 = e(E_1) = e(e_1(I)) = I$$

y

$$E_1E = e_1(E) = e_1(e(I)) = I$$

con lo que E es inversible y $E_1 = E^{-1}$. $\blacksquare$

Ejemplo 14.

(a) $$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

(b) $$\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -c \\ 0 & 1 \end{bmatrix}$$

(c) $$\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix}$$

(d) Cuando $c \neq 0$,

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} c^{-1} & 0 \\ 0 & 1 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & c^{-1} \end{bmatrix}.$$

Teorema 12. *Si A es una matriz $n \times n$, los siguientes enunciados son equivalentes.*

(i) *A es inversible.*
(ii) *A es equivalente por filas a la matriz identidad $n \times n$*
(iii) *A es un producto de matrices elementales.*

Demostración. Sea R una matriz escalón reducida por filas, equivalente por filas a A. Por el Teorema 9 (o su corolario),

$$R = E_k \cdots E_2E_1A$$

donde $E_1, \ldots, E_k$ son matrices elementales. Cada E_j es inversible, y así

$$A = E_1^{-1} \cdots E_k^{-1}R.$$

Como el producto de matrices inversibles es inversible, se ve que A es inversible si, y solo si, R es inversible. Como R es una matriz (cuadrada) escalón reducida por filas, R es inversible si, y solo si, toda fila de R contiene un elemento no nulo, esto es, si, y solo si, $R = I$. Hemos visto que A es inversible si, y solo si, $R = I$, y si $R = I$ entonces $A = E_k^{-1}, \ldots, E_1^{-1}$. Se ve ahora que (i), (ii) y (iii) son afirmaciones equivalentes respecto de A. ▮

Corolario. *Si A es una matriz inversible $n \times n$ y si una sucesión de operaciones elementales de fila reduce A a la identidad, entonces la misma sucesión de operaciones, cuando se aplica a I, da A^{-1}.*

Corolario. *Sean A y B dos matrices $m \times n$. Entonces B es equivalente por filas a A si, y solo si, $B = PA$, donde P es una matriz $m \times m$ inversible.*

Teorema 13. *Para una matriz $n \times n$, las siguientes afirmaciones son equivalentes.*

(i) *A es inversible.*

(ii) *El sistema homogéneo $AX = 0$ tiene solo la solución trivial $X = 0$.*

(iii) *El sistema de ecuaciones $AX = Y$ tiene una solución X para cada matriz $n \times 1$, Y.*

Demostración. De acuerdo con el Teorema 7, la condición (ii) es equivalente al hecho de que A es equivalente por filas a la matriz identidad. Por el Teorema 12, (i) y (ii) son, por tanto, equivalentes. Si A es inversible, la solución de $AX = Y$ es $X = A^{-1}Y$. Recíprocamente, supóngase que $AX = Y$ tiene una solución para cada Y dada. Sea R una matriz reducida por filas que sea equivalente por filas a A. Se debe demostrar que $R = I$. Esto equivale a demostrar que la última fila de R no es 0 (idénticamente). Sea

$$E = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Si el sistema $RX = E$ puede resolverse para X, la última fila de R no puede ser 0. Se sabe que $R = PA$, donde P es inversible. Luego $RX = E$ si, y solo si, $AX = P^{-1}E$. De acuerdo con (iii), el último sistema tiene una solución. ▮

Corolario. *Una matriz cuadrada que tiene una inversa a la izquierda o a la derecha es inversible.*

Demostración. Sea A una matriz $m \times n$. Supóngase que A tiene una inversa a la izquierda, es decir, una matriz B tal que $BA = I$. Entonces $AX = 0$ tiene solo la solución trivial, porque $X = IX = B(AX)$. Luego A es inversible. Por otro lado, supóngase que A tiene una inversa a la derecha, es decir, una matriz C tal que $AC = I$. Entonces C tiene una inversa a la izquierda y es, por tanto, inversible. Se sigue entonces que $A = C^{-1}$ y así A es inversible con inversa C. ▮

Corolario. *Sea $A = A_1A_2 \cdots A_k$, donde $A_1, \ldots, A_k$ son matrices (cuadradas) $n \times n$. Entonces A es inversible si, y solo si, cada A_j es inversible.*

Demostración. Se ha visto que el producto de dos matrices inversibles es inversible. De lo cual es fácil deducir que si cada A_j es inversible entonces A es inversible.

Supóngase ahora que A es inversible. Se demostrará primero que A_k es inversible. Supóngase que X es una matriz $n \times 1$ y que $A_kX = 0$. Entonces $AX = (A_1, \ldots, A_{k-1})A_kX = 0$. Como A es inversible se debe tener que $X = 0$. El sistema de ecuaciones $A_kX = 0$ no tiene, pues, solución no trivial, y así A_k es inversible. Pero ahora $A_1 \cdots A_{k-1} = AA_k^{-1}$ es inversible. Por lo dicho antes, A_{k-1} es inversible. Continuando en esta forma se concluye que todo A_j es inversible. ▮

Quisiéramos hacer ahora un comentario final respecto a la solución de ecuaciones lineales. Supóngase que A es una matriz $m \times n$ y que deseamos resolver el sistema de ecuaciones $AX = Y$. Si R es una matriz escalón reducida por filas, equivalente por filas a A, entonces $R = PA$, donde P es una matriz $m \times m$ inversible. Las soluciones del sistema $AX = Y$ son exactamente las mismas que las soluciones del sistema $RX = PY(=Z)$. En la práctica no es mucho más difícil hallar la matriz P que reducir por filas A a R. En efecto, supóngase que se forma la matriz aumentada A' del sistema $AX = Y$, con escalares arbitrarios $y_1, \ldots, y_m$ en su última columna. Si entonces se efectúa sobre A' una sucesión de operaciones elementales de fila que lleve de A a R, se habrá aclarado qué es la matriz P. (El lector deberá remitirse al Ejemplo 9, donde se llevó adelante esencialmente este proceso.) En particular si A es una matriz cuadrada este proceso deja en claro cuándo A es o no inversible, y si A es inversible cuál es la inversa P. Como ya se ha dado la parte principal de un ejemplo de tal cálculo, solo se dará ahora un ejemplo de 2×2.

Ejemplo 15. Supóngase que F es el cuerpo de los números racionales y

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}.$$

Así

$$\begin{bmatrix} 2 & -1 & y_1 \\ 1 & 3 & y_2 \end{bmatrix} \xrightarrow{(3)} \begin{bmatrix} 1 & 3 & y_2 \\ 2 & -1 & y_1 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & 3 & y_2 \\ 0 & -7 & y_1 - 2y_2 \end{bmatrix} \xrightarrow{(1)}$$

$$\begin{bmatrix} 1 & 3 & y_2 \\ 0 & 1 & \frac{1}{7}(2y_2 - y_1) \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & 0 & \frac{1}{7}(y_2 + 3y_1) \\ 0 & 1 & \frac{1}{7}(2y_2 - y_1) \end{bmatrix}$$

de lo que resulta claro que A es inversible y

$$A^{-1} = \begin{bmatrix} \frac{3}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{2}{7} \end{bmatrix}.$$

Puede parecer engorroso seguir escribiendo los escalares arbitrarios y_1, y_2, . . . en el cálculo de las inversas. Algunos encuentran más fácil proceder con dos sucesiones de matrices, una que describa la reducción de A a la identidad, y otra que registre el efecto de la misma sucesión de operaciones partiendo de la identidad. El lector juzgará por sí mismo cuál es la forma que le resulta más cómoda para contabilizar los cálculos.

Ejemplo 16. Hallar la inversa de

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}.$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & \frac{1}{12} & \frac{4}{45} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{3} & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & 0 & \frac{1}{180} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ \frac{1}{6} & -1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -6 & 12 & 0 \\ 30 & -180 & 180 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} -9 & 60 & -60 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}.$$

El lector se habrá percatado que se ha hecho un largo estudio sobre las filas de una matriz y no se ha dicho casi nada de las columnas. Se ha concentrado la atención en las filas por parecer ello más natural desde el punto de vista de las ecuaciones lineales. Dado que, obviamente, no hay ningún privilegio con respecto a las filas, lo tratado en la última sección pudo haberse hecho usando columnas en vez de filas. Si se define una operación elemental de columna y equivalencia por columna, es claro que toda matriz $m \times n$ será equivalente

por columnas a una matriz «escalón reducida por columnas». Así cada operación elemental de columna será de la forma $A \longrightarrow AE$, donde E es una matriz elemental $n \times n$, y así sucesivamente.

Ejercicios

1. Sea

$$A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ -1 & 0 & 3 & 5 \\ 1 & -2 & 1 & 1 \end{bmatrix}.$$

Hallar una matriz escalón reducida por filas R que sea equivalente a A, y una matriz inversible 3×3, P, tal que $R = PA$.

2. Repetir el Ejercicio 1, pero con

$$A = \begin{bmatrix} 2 & 0 & i \\ 1 & -3 & -i \\ i & 1 & 1 \end{bmatrix}.$$

3. Para cada una de las dos matrices

$$\begin{bmatrix} 2 & 5 & -1 \\ 4 & -1 & 2 \\ 6 & 4 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 & 2 \\ 3 & 2 & 4 \\ 0 & 1 & -2 \end{bmatrix}$$

emplear operaciones elementales de fila para determinar cuándo es inversible y encontrar la inversa en caso que lo sea.

4. Sea

$$A = \begin{bmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 0 & 1 & 5 \end{bmatrix}.$$

¿Para qué X existe un escalar c tal que $AX = cX$?

5. Determinar si

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 3 & 4 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

es inversible y hallar A^{-1} si existe.

6. Supóngase que A es una matriz 2×1 y que B es una matriz 1×2. Demostrar que $C = AB$ no es inversible.

7. Sea A una matriz (cuadrada) $n \times n$. Demostrar las siguientes dos afirmaciones:

(a) Si A es inversible y $AB = 0$ para alguna matriz $n \times n$, B, entonces $B = 0$.

(b) Si A no es inversible, entonces existe una matriz $n \times n$, B, tal que $AB = 0$, pero $B \neq 0$.

8. Sea

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Demostrar, usando operaciones elementales de fila, que A es inversible si, y solo si, $(ad - bc) \neq 0$.

9. Una matriz $n \times n$, A, se llama **triangular superior** si $A_{ij} = 0$ para $i > j$; esto es, si todo elemento por debajo de la diagonal principal es 0. Demostrar que una matriz (cuadrada) triangular superior es inversible si, y solo si, cada elemento de su diagonal principal es diferente de 0.

10. Demostrar la siguiente generalización del Ejercicio 6. Si A es una matriz $m \times n$, B es una matriz $n \times m$ y $n < m$, entonces AB no es inversible.

11. Sea A una matriz $m \times n$. Hacer ver que por medio de un número finito de operaciones elementales de fila y/o de columna se puede pasar de A a la matriz R que es simultáneamente «escalón reducida por filas» y «escalón reducida por columnas»; es decir, $R_{ij} = 0$ si $i \neq j$, $R_{ii} = 1$, $1 \leq i \leq r$, $R_{ii} = 0$ si $i > r$. Demostrar que $R = PAQ$, donde P es una matriz inversible $m \times m$ y Q es una matriz inversible $n \times n$.

12. El resultado del Ejemplo 16 sugiere que tal vez la matriz

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2n-1} \end{bmatrix}$$

es inversible y A^{-1} tiene elementos enteros. ¿Se puede demostrar esto?

2. *Espacios vectoriales*

2.1. Espacios vectoriales

En varias partes de la matemática se presentan conjuntos donde tiene sentido y resulta interesante considerar las «combinaciones lineales» de los elementos de dichos conjuntos. Por ejemplo, en los sistemas de ecuaciones lineales, se encontró natural el considerar combinaciones lineales de las filas de una matriz. Es probable que el lector ya haya estudiado cálculo y haya tenido que ver con combinaciones lineales de funciones; sobre todo si ha estudiado ecuaciones diferenciales. Tal vez el lector haya tenido experiencia con vectores en el espacio euclidiano tridimensional, y en particular con combinaciones lineales de tales vectores.

Hablando en forma simple, el álgebra lineal es aquella rama de la matemática que trata de las propiedades comunes de los sistemas algebraicos, que constan de un conjunto más una noción razonable de «combinación lineal» de los elementos del conjunto. En esta sección se definirán los objetos matemáticos que la experiencia ha mostrado ser las más útiles abstracciones de este tipo de sistemas algebraicos.

Definición. *Un* **espacio vectorial** (*o espacio lineal*) *consta de lo siguiente*:

1. *un cuerpo F de escalares*;
2. *un conjunto V de objetos llamados vectores*;
3. *una regla* (*u operación*) *llamada adición, que asocia a cada par de vectores α, β de V un vector $\alpha + \beta$ de V, que se llama suma de α y β, de tal modo que*:
   (a) *la adición es conmutativa*, $\alpha + \beta = \beta + \alpha$;
   (b) *la adición es asociativa*, $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$;

(c) *existe un único vector 0 de V, llamado vector nulo, tal que* $\alpha + 0 = \alpha$, *para todo* α *de* V;

(d) *para cada vector* α *de* V, *existe un único vector* $-\alpha$ *de* V, *tal que* $\alpha + (-\alpha) = 0$;

4. *una regla (u operación), llamada multiplicación escalar, que asocia a cada escalar* c *de* F *y cada vector* α *de* V *a un vector* $c\alpha$ *en* V, *llamado producto de* c *y* α, *de tal modo que*:

(a) $1\alpha = \alpha$ *para todo* α *de* V;
(b) $(c_1c_2)\alpha = c_1(c_2\alpha)$;
(c) $c(\alpha + \beta) = c\alpha + c\beta$;
(d) $(c_1 + c_2)\alpha = c_1\alpha + c_2\alpha$.

Es importante observar, como la definición establece, que un espacio vectorial es un objeto compuesto que consta de un cuerpo, de un conjunto de «vectores» y de dos operaciones con ciertas propiedades especiales. El mismo conjunto de vectores puede ser parte de distintos espacios vectoriales (véase Ejemplo 5, más adelante). Cuando no hay posibilidad de confusión, se hará referencia simplemente al espacio vectorial V, y cuando se desee especificar el cuerpo, se dirá que V es un **espacio vectorial sobre el cuerpo** F. El nombre «vector» se da a los elementos del conjunto V más bien por conveniencia. El origen del nombre se encontrará en el Ejemplo 1, pero no se debe dar mucha importancia al nombre mismo, ya que la variedad de objetos que pueden ser vectores en V puede no parecerse gran cosa a algún concepto de vector que el lector ya tenga. Se tratará de ilustrar esta variedad en una serie de ejercicios que será ampliada considerablemente cuando se comience el estudio sistemático de los espacios vectoriales.

Ejemplo 1. **El espacio de *n*-tuples, F^n.** Sea F cualquier cuerpo y sea V el conjunto de todos los n-tuples $\alpha = (x_1, x_2, \ldots, x_n)$ de escalares x_i de F. Si $\beta = (y_1, y_2, \ldots, y_n)$ con y_i de F, la suma de α y β se define por

$$\alpha + \beta = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n). \tag{2-1}$$

El producto de un escalar c y el vector α se define por

$$c\alpha = (cx_1, cx_2, \ldots, cx_n). \tag{2-2}$$

Que esta adición vectorial y multiplicación escalar cumplen las condiciones (3) y (4) es fácil de verificar, usando las propiedades semejantes de la adición y multiplicación de los elementos de F.

Ejemplo 2. **El espacio de matrices $m \times n$, $F^{m\times n}$.** Sea F cualquier cuerpo y sean m y n enteros positivos. Sea $F^{m\times n}$ el conjunto de todas las matrices $m \times n$ sobre el cuerpo F. La suma de dos vectores A y B en $F^{m\times n}$ se define por

$$(A + B)_{ij} = A_{ij} + B_{ij}. \tag{2-3}$$

El producto de un escalar c y de la matriz A se define por

(2-4) $$(cA)_{ij} = cA_{ij}.$$

Obsérvese que $F^{1 \times n} = F^n$.

Ejemplo 3. **El espacio de funciones de un conjunto en un cuerpo.** Sea F cualquier cuerpo y sea S cualquier conjunto no vacío. Sea V el conjunto de todas las funciones de S en F. La suma de dos vectores f y g de V es el vector $f + g$; es decir, la función de S en F definida por

(2-5) $$(f + g)(s) = f(s) + g(s).$$

El producto del escalar c y la función f es la función cf definida por

(2-6) $$(cf)(s) = cf(s).$$

Los ejemplos anteriores son un caso particular de este último. En efecto, un n-tuple de elementos de F puede considerarse como una función del conjunto S de los enteros $1, \ldots, n$ de F. En forma análoga, una matriz $m \times n$ sobre el cuerpo F es una función del conjunto S de los pares de enteros (i, j), $1 \leq i \leq m$, $1 \leq j \leq n$, en el cuerpo F. Para este tercer ejemplo se indica cómo se puede verificar que las operaciones definidas satisfacen las condiciones (3) y (4). Para la adición vectorial:

(a) Como la adición en F es conmutativa,

$$f(s) + g(s) = g(s) + f(s)$$

para todo s de S, luego las funciones $f + g$ y $g + f$ son idénticas.

(b) Como la adición en F es asociativa,

$$f(s) + [g(s) + h(s)] = [f(s) + g(s)] + h(s)$$

para todo s, luego $f + (g + h)$ es la misma función que $(f + g) + h$.

(c) El único vector nulo es la función cero, que asigna a cada elemento de S el escalar 0 de F.

(d) Para todo f de V, $(-f)$ es la función dada por

$$(-f)(s) = -f(s).$$

El lector encontrará fácil verificar que la multiplicación escalar satisface las condiciones de (4), razonando como se hizo para la adición vectorial.

Ejemplo 4. **El espacio de las funciones polinomios sobre el cuerpo F.** Sea F un cuerpo y sea V el conjunto de todas las funciones f de F en F definidas en la forma

(2-7) $$f(x) = c_0 + c_1x + \cdots + c_nx^n$$

donde $c_0, c_1, \ldots, c_n$ son escalares fijos de F (independiente de x). Una función de este tipo se llama **función polinomio sobre F.** Sean la adición y la multiplicación escalar definidas como en el Ejemplo 3. Se debe observar que si f y g

son funciones polinomios y c está en F, entonces $f + g$ y cf son también funciones polinomios.

Ejemplo 5. El cuerpo C de los números complejos puede considerarse como un espacio vectorial sobre el cuerpo R de los números reales. En forma más general, sea F el cuerpo de los números reales y sea V el conjunto de los n-tuples $\alpha = (x_1, \ldots, x_n)$, donde $x_1, \ldots, x_n$ son números *complejos*. Se define la adición vectorial y la multiplicación escalar por (2-1) y (2-2), como en el Ejemplo 1. De este modo se obtiene un espacio vectorial sobre el cuerpo R que es muy diferente del espacio C^n y del espacio R^n.

Hay unos pocos hechos simples que se desprenden, casi inmediatamente, de la definición de espacio vectorial, y procederemos a derivarlos. Si c es un escalar y 0 es el vector nulo, entonces por 3(c) y 4(c)

$$c0 = c(0 + 0) = c0 + c0.$$

Sumando $-(c0)$ y por 3(d), se obtiene

$$c0 = 0. \tag{2-8}$$

Análogamente, para el escalar 0 y cualquier vector α se tiene que

$$0\alpha = 0. \tag{2-9}$$

Si c es un escalar no nulo y α un vector tal que $c\alpha = 0$, entonces por (2-8), $c^{-1}(c\alpha) = 0$. Pero

$$c^{-1}(c\alpha) = (c^{-1}c)\alpha = 1\alpha = \alpha$$

luego, $\alpha = 0$. Así se ve que si c es un escalar y α un vector tal que $c\alpha = 0$, entonces c es el escalar cero o α es el vector nulo.

Si α es cualquier vector de V, entonces

$$0 = 0\alpha = (1 - 1)\alpha = 1\alpha + (-1)\alpha = \alpha + (-1)\alpha$$

de lo que se sigue que

$$(-1)\alpha = -\alpha. \tag{2-10}$$

Finalmente, las propiedades asociativa y conmutativa de la adición vectorial implican que la suma de varios vectores es independiente de cómo se combinen estos vectores y de cómo se asocien. Por ejemplo, si $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ son vectores de V, entonces

$$(\alpha_1 + \alpha_2) + (\alpha_3 + \alpha_4) = [\alpha_2 + (\alpha_1 + \alpha_3)] + \alpha_4$$

y tal suma puede ser escrita, sin confusión,

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4.$$

Definición. *Un vector β de V se dice* **combinación lineal** *de los vectores $\alpha_1, \ldots, \alpha_n$ en V, si existen escalares $c_1, \ldots, c_n$ de F tales que*

$$\beta = c_1\alpha_1 + \cdots + c_n\alpha_n = \sum_{i=1}^{n} c_i\alpha_i.$$

Otras extensiones de la propiedad asociativa de la adición vectorial y las propiedades distributivas 4(c) y 4(d) de la multiplicación escalar se aplican a las combinaciones lineales:

$$\sum_{i=1}^{n} c_i\alpha_i + \sum_{i=1}^{n} d_i\alpha_i = \sum_{i=1}^{n} (c_i + d_i)\alpha_i$$

$$c \sum_{i=1}^{n} c_i\alpha_i = \sum_{i=1}^{n} (cc_i)\alpha_i.$$

Ciertas partes del álgebra lineal están íntimamente relacionadas con la geometría. La misma palabra «espacio» sugiere algo geométrico, como lo hace el vocablo «vector» para muchos. Cuando se avance en el estudio de los espacios vectoriales, el lector observará que mucha de la terminología tiene una connotación geométrica. Para concluir esta sección introductoria sobre espacios vectoriales, se considerará la relación de los espacios vectoriales con la geometría, hasta un grado que indicará al menos el origen del nombre «espacio vectorial». Este será un breve análisis intuitivo.

Considérese el espacio vectorial R^3. En geometría analítica se identifican las ternas (x_1, x_2, x_3) de números reales con los puntos del espacio euclidiano tridimensional. En tal contexto, un vector se suele definir como un segmento dirigido PQ, del punto P del espacio a otro punto Q. Ello equivale a una formulación cuidadosa de la idea de «flecha» de P a Q. Tal como son usados los vectores, se ha tenido en mente que queden definidos por su longitud y dirección. Así, se deben identificar dos segmentos dirigidos si tienen la misma longitud y la misma dirección.

El segmento dirigido PQ, del punto $P = (x_1, x_2, x_3)$ al punto $Q = (y_1, y_2, y_3)$, tiene la misma longitud y dirección que el segmento dirigido del origen $0 = (0, 0, 0)$ al punto $(y_1 - x_1, y_2 - x_2, y_3 - x_3)$. Además, este es el único segmento que partiendo del origen tiene la misma longitud y dirección que PQ. Con lo que, si se conviene en operar solo con vectores aplicados al origen, hay exactamente un vector asociado con cada longitud y dirección dadas.

El vector OP, del origen a $P = (x_1, x_2, x_3)$ está completamente determinado por P, y es por ello posible identificar este vector con el punto P. En la definición del espacio vectorial R^3, los vectores se definen simplemente como las ternas (x_1, x_2, x_3).

Dados los puntos $P = (x_1, x_2, x_3)$ y $Q = (y_1, y_2, y_3)$, la definición de suma de los vectores OP y OQ puede hacerse geométricamente. Si los vectores no son paralelos, entonces los segmentos OP y OQ determinan un plano y estos segmentos son dos de los lados de un paralelogramo en aquel plano (véase Figura 1). La diagonal de este paralelogramo, que va de 0 al punto S, define la suma de OP y OQ como el vector OS. Las coordenadas del punto S son $(x_1 + y_1, x_2 + y_2, x_3 + y_3)$, con lo que esta definición geométrica de la adición vectorial es equivalente a la definición algebraica del Ejemplo 1.

La multiplicación escalar tiene una interpretación geométrica simple. Si c es un número real, entonces el producto de c por el vector OP es el vector que parte del origen de longitud $|c|$ veces la longitud de OP y dirección que coin-

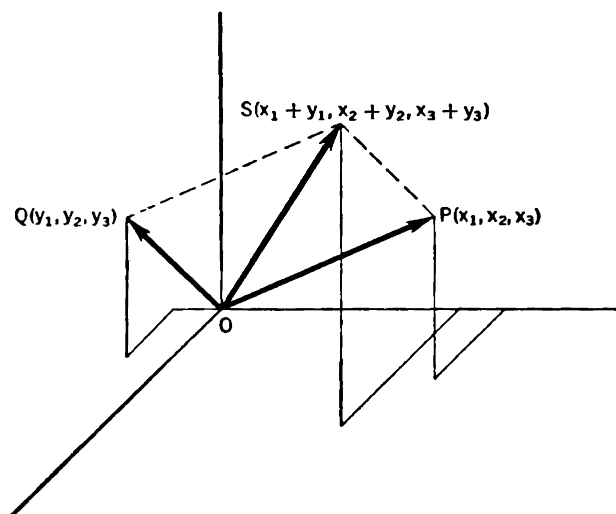


FIGURA 1

cide con la de OP si $c > 0$, y que es opuesta a la dirección de OP si $c < 0$. Esta multiplicación escalar da justamente el vector OT donde $T = (cx_1, cx_2, cx_3)$, y es, por tanto, compatible con la definición algebraica dada en R^3.

De vez en cuando el lector encontrará probablemente provechoso «pensar geométricamente» en espacios vectoriales, eso es, trazar gráficos que lo ayuden a ilustrar y motivar alguna de las ideas. Ciertamente, deberá hacerlo. Sin embargo, al hacer tales ilustraciones, debe tenerse presente que, por tratar los espacios vectoriales como sistemas algebraicos, todas las demostraciones que se hagan deben ser de naturaleza algebraica.

Ejercicios

1. Si F es un cuerpo, verificar que F^n (como se definió en el Ejemplo 1) es un espacio vectorial sobre el cuerpo F.

2. Si V es un espacio vectorial sobre el cuerpo F, verificar que

$$(\alpha_1 + \alpha_2) + (\alpha_3 + \alpha_4) = [\alpha_2 + (\alpha_3 + \alpha_1)] + \alpha_4$$

para todos los vectores α_1, α_2, α_3 y α_4 de V.

3. Si C es el cuerpo de los complejos, ¿qué vectores de C^3 son combinaciones lineales de $(1, 0, -1)$, $(0, 1, 1)$ y $(1, 1, 1)$?

4. Sea V el conjunto de los pares (x, y) de números reales, y sea F el cuerpo de los números reales. Se define

$$(x, y) + (x_1, y_1) = (x + x_1, y + y_1)$$
$$c(x, y) = (cx, y).$$

¿Es V, con estas operaciones, un espacio vectorial sobre el cuerpo de los números reales?

5. En R^n se definen dos operaciones

$$\alpha \oplus \beta = \alpha - \beta$$
$$c \cdot \alpha = -c\alpha.$$

Las operaciones del segundo miembro son las usuales. ¿Qué axiomas de espacio vectorial se cumplen para $(R^n, \oplus, \cdot)$?

6. Sea V el conjunto de todas las funciones que tienen valor complejo sobre el eje real, tales que (para todo t de R)

$$f(-t) = \overline{f(t)}.$$

La barra indica conjugación compleja. Demostrar que V, con las operaciones

$$(f + g)(t) = f(t) + g(t)$$
$$(cf)(t) = cf(t)$$

es un espacio vectorial sobre el cuerpo de los números reales. Dar un ejemplo de una función en V que no toma valores reales.

7. Sea V el conjunto de pares (x, y) de números reales y sea F el cuerpo de los números reales. Se define

$$(x, y) + (x_1, y_1) = (x + x_1, 0)$$
$$c(x, y) = (cx, 0).$$

¿Es V, con estas operaciones, un espacio vectorial?

2.2. *Subespacios*

En esta sección se introducirán algunos de los conceptos básicos en el estudio de los espacios vectoriales.

Definición. *Sea V un espacio vectorial sobre el cuerpo F. Un* **subespacio** *de V es un subconjunto W de V que, con las operaciones de adición vectorial y multiplicación escalar sobre V, es el mismo un espacio vectorial sobre F.*

Una comprobación directa de los axiomas para un espacio vectorial, demuestra que el subconjunto W de V es un subespacio si, para todos los α y β de W, el vector $\alpha + \beta$ está también en W; el vector 0 está en W; para todo α de W, el vector $(-\alpha)$ está en W; para todo α de W y todo escalar c, el vector $c\alpha$ está en W. La conmutatividad y asociatividad de la adición vectorial y las propiedades (4)(a), (b), (c) y (d) de la multiplicación escalar no necesitan ser comprobadas, ya que éstas son propiedades de las operaciones en V. Se pueden simplificar aún más las cosas.

Teorema 1. *Un subconjunto no vacío W de V es un subespacio de V si, y solo si, para todo par de vectores α, β de W y todo escalar c de F, el vector $c\alpha + \beta$ está en W.*

Demostración. Supóngase que W sea un subconjunto no vacío de V tal

que $c\alpha + \beta$ pertenezca a W para todos los vectores α, β de W y todos los escalares c de F. Como W no es vacío, existe un vector ρ en W, y, por tanto, $(-1)\rho + \rho = 0$ está en W. Ahora bien, si α es cualquier vector de W y c cualquier escalar, el vector $c\alpha = c\alpha + 0$ está en W. En particular, $(-1)\alpha = -\alpha$ está en W. Finalmente, si α y β están en W, entonces $\alpha + \beta = 1\alpha + \beta$ está en W. Así, W es un subespacio de V.

Recíprocamente, si W es un subespacio de V, α y β están en W y c es un escalar, ciertamente $c\alpha + \beta$ está en W. ▌

Algunos prefieren usar la propiedad $c\alpha + \beta$ del Teorema 1 como definición de un subespacio, lo que es apenas diferente. Lo importante es que, si W es un subconjunto no vacío de V tal que $c\alpha + \beta$ está en W para todos los α, β de W y todo c de F, entonces W (con las operaciones heredadas de V) sea un espacio vectorial. Esto da lugar a muchos nuevos ejemplos de espacios vectoriales.

Ejemplo 6.

(a) Si V es cualquier espacio vectorial, V es un subespacio de V; el subconjunto que consta solo del vector nulo es un subespacio vectorial de V, llamado **subespacio nulo** de V*.

(b) En F^n, el conjunto de los n-tuples $(x_1, \ldots, x_n)$ con $x_1 = 0$ es un subespacio, pero el conjunto de los n-tuples con $x_1 = 1 + x_2$ no es un subespacio $(n \geq 2)$.

(c) El espacio de las funciones polinomios sobre el cuerpo F es un subespacio del espacio de todas las funciones de F en F.

(d) Una matriz (cuadrada) $n \times n$, sobre el cuerpo F es **simétrica** si $A_{ij} = A_{ji}$ para todo i y j. Las matrices simétricas forman un subespacio del espacio de las matrices $n \times n$ sobre F.

(e) Una matriz (cuadrada) $n \times n$, A, sobre el cuerpo C de los números complejos es **hermítica (o autoadjunta)** si

$$A_{jk} = \overline{A_{kj}}$$

para todo j, k, donde la barra indica conjugación compleja. Una matriz 2×2 es hermítica si, y solo si, tiene la forma

$$\begin{bmatrix} z & x + iy \\ x - iy, & w \end{bmatrix}$$

donde x, y, z y w son números reales. El conjunto de todas las matrices hermíticas *no* es un subespacio del espacio de todas las matrices $n \times n$ sobre C. En efecto, si A es hermítica, sus elementos en la diagonal $A_{11}, A_{22}, \ldots$, son números reales, pero los elementos diagonales de iA, en general, no son reales. Por otro lado, es fácil ver que el conjunto de las matrices hermíticas $n \times n$ es un espacio vectorial sobre el cuerpo de los números reales (con las operaciones usuales).

* Nota del traductor: Estos subespacios se llaman comúnmente los **subespacios triviales de** V.

Ejemplo 7. **El espacio solución de un sistema homogéneo de ecuaciones lineales.** Sea A una matriz $m \times n$ sobre F. Entonces el conjunto de todas las matrices (columna) $n \times 1$, X, sobre F tal que $AX = 0$ es un subespacio del espacio de todas las matrices $n \times 1$ sobre F. Para demostrar esto se necesita probar que $A(cX + Y) = 0$ si $AX = 0$, $AY = 0$ y c un escalar arbitrario de F. Esto se desprende inmediatamente del siguiente hecho general.

Lema. *Si A es una matriz $m \times n$ sobre F, y B, C son matrices $n \times p$ sobre F, entonces*

(2-11) $$A(dB + C) = d(AB) + AC$$

para todo escalar d de F.

Demostración.

$$\begin{aligned}[A(dB + C)]_{ij} &= \sum_k A_{ik}(dB + C)_{kj} \\ &= \sum_k (dA_{ik}B_{kj} + A_{ik}C_{kj}) \\ &= d\sum_k A_{ik}B_{kj} + \sum_k A_{ik}C_{kj} \\ &= d(AB)_{ij} + (AC)_{ij} \\ &= [d(AB) + AC]_{ij}. \quad \blacksquare\end{aligned}$$

En forma semejante se puede ver que $(dB + C)A = d(BA) + CA$, si la suma y el producto de las matrices están definidos.

Teorema 2. *Sea V un espacio vectorial sobre el cuerpo F. La intersección de cualquier colección de subespacios de V es un subespacio de V.*

Demostración. Sea $\{W_a\}$ una colección de subespacios de V, y sea $W = \bigcap_a W_a$ su intersección. Recuérdese que W está definido como el conjunto de todos los elementos pertenecientes a cada W_a (véase Apéndice). Dado que todo W_a es un subespacio, cada uno contiene el vector nulo. Así que el vector nulo está en la intersección W, y W no es vacío. Sean α y β vectores de W y sea c un escalar. Por definición de W ambos, α y β, pertenecen a cada W_a, y por ser cada W_a un subespacio el vector $(c\alpha + \beta)$ está en cada W_a. Así $(c\alpha + \beta)$ está también en W. Por el Teorema 1, W es un subespacio de V. $\blacksquare$

Del Teorema 2 se deduce que si S es cualquier colección de vectores de V, entonces existe un subespacio mínimo de V que contiene a S; esto es, un subespacio que contiene a S y que está contenido en cada uno de los otros subespacios que contienen a S.

Definición. *Sea S un conjunto de vectores de un espacio vectorial V. El* **subespacio generado** *por S se define como intersección W de todos los subespacios de V que contienen a S. Cuando S es un conjunto finito de vectores, $S = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ se dice simplemente que W es el* **subespacio generado por los vectores** *$\alpha_1, \alpha_2, \ldots, \alpha_n$.*

Teorema 3. *El subespacio generado por un subconjunto S no vacío de un espacio vectorial V es el conjunto de todas las combinaciones lineales de los vectores de S.*

Demostración. Sea W el subespacio generado por S. Entonces toda combinación lineal

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_m\alpha_m$$

de vectores $\alpha_1, \alpha_2, \ldots, \alpha_m$ de S pertenece evidentemente a W. Así que W contiene el conjunto L de todas las combinaciones lineales de vectores de S. El conjunto L, por otra parte, contiene a S y no es, pues, vacío. Si α, β pertenecen a L, entonces α es una combinación lineal,

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_m\alpha_m$$

de vectores α_i de S, y β es una combinación lineal,

$$\beta = y_1\beta_1 + y_2\beta_2 + \cdots + y_n\beta_n$$

de vectores β_j en S. Para cada escalar c,

$$c\alpha + \beta = \sum_{i=1}^{m} (cx_i)\alpha_i + \sum_{j=1}^{n} y_j\beta_j.$$

Luego $c\alpha + \beta$ pertenece a L. Con lo que L es un subespacio de V.

Se ha demostrado que L es un subespacio de V que contiene a S, y también que todo subespacio que contiene a S contiene a L. Se sigue que L es la intersección de todos los subespacios que contienen a S; es decir, que L es el subespacio generado por el conjunto S. ∎

Definición. *Si $S_1, S_2, \ldots, S_k$ son subconjuntos de un espacio vectorial V, el conjunto de todas las sumas*

$$\alpha_1 + \alpha_2 + \cdots + \alpha_k$$

de vectores α_i de S_i se llama **suma** *de los subconjuntos $S_1, S_2, \ldots, S_k$ y se representa por*

$$S_1 + S_2 + \cdots + S_k$$

o por

$$\sum_{i=1}^{k} S_i.$$

Si $W_1, W_2, \ldots, W_k$ son subespacios de V, entonces la suma

$$W = W_1 + W_2 + \cdots + W_k$$

como es fácil ver, es un subespacio de V que contiene cada uno de los subespacios W_i. De esto se sigue, como en la demostración del Teorema 3, que W es el subespacio generado por la unión de $W_1, W_2, \ldots, W_k$.

Ejemplo 8. Sea F un subcuerpo del cuerpo C de los números complejos. Supóngase que

$$\alpha_1 = (1, 2, 0, 3, 0)$$
$$\alpha_2 = (0, 0, 1, 4, 0)$$
$$\alpha_3 = (0, 0, 0, 0, 1).$$

Por el Teorema 3, un vector α está en el subespacio W de F^5 generado por α_1, α_2, α_3 si, y solo si, existen escalares c_1, c_2, c_3 en F, tales que

$$\alpha = c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3.$$

Así, W consta de todos los vectores de la forma

$$\alpha = (c_1, 2c_1, c_2, 3c_1 + 4c_2, c_3)$$

donde c_1, c_2, c_3 son escalares arbitrarios de F. En forma alternativa, W puede ser escrito como el conjunto de todos los 5-tuples

$$\alpha = (x_1, x_2, x_3, x_4, x_5)$$

con x_i en F, tal que

$$x_2 = 2x_1$$
$$x_4 = 3x_1 + 4x_3.$$

Así $(-3, -6, 1, -5, 2)$ está en W, mientras que $(2, 4, 6, 7, 8)$ no.

Ejemplo 9. Sea F un subcuerpo del cuerpo de los números complejos, y sea V el espacio vectorial de todas las matrices 2×2 sobre F. Sea W_1 el subconjunto de V que consta de todas las matrices de la forma

$$\begin{bmatrix} x & y \\ z & 0 \end{bmatrix}$$

donde x, y, z son escalares arbitrarios de F. Por último, sea W_2 el subconjunto de V que consta de todas las matrices de la forma

$$\begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}$$

donde x e y son escalares arbitrarios de F. Entonces W_1 y W_2 son subespacios de V. También

$$V = W_1 + W_2$$

porque

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix}.$$

El subespacio $W_1 \cap W_2$ consta de todas las matrices de la forma

$$\begin{bmatrix} x & 0 \\ 0 & 0 \end{bmatrix}.$$

Ejemplo 10. Sea A una matriz $m \times n$ sobre el cuerpo F. Los **vectores fila** de A son los vectores de F^n dados por $\alpha_i = (A_{i1}, \ldots, A_{in})$, $i = 1, \ldots, m$. El subespacio de F^n generado por los vectores fila se llama el **espacio de filas** de A. El subespacio considerado en el Ejemplo 8 es el espacio de filas de la matriz

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Es también el espacio de filas de la matriz

$$B = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -4 & -8 & 1 & -8 & 0 \end{bmatrix}.$$

Ejemplo 11. Sea V el espacio de todas las funciones polinomios sobre F. Sea S el subconjunto de V que consta de las funciones polinomios $f_0, f_1, f_2, \ldots$ definido por

$$f_n(x) = x^n, \qquad n = 0, 1, 2, \ldots.$$

Entonces V es el subespacio generado por el conjunto S.

Ejercicios

1. ¿Cuáles de los siguientes conjuntos de vectores $\alpha = (a_1, \ldots, a_n)$ de R^n son subespacios de R^n $(n \geq 3)$?

(a) todos los α tal que $a_1 \geq 0$;
(b) todos los α tal que $a_1 + 3a_2 = a_3$;
(c) todos los α tal que $a_2 = a_1$;
(d) todos los α tal que $a_1 a_2 = 0$;
(e) todos los α tal que a_2 es racional.

2. Sea V el espacio vectorial (real) de todas las funciones f de R en R. ¿Cuál de los siguientes conjuntos de funciones son subespacios de V?

(a) todas las f tales que $f(x^2) = f(x)^2$;
(b) todas las f tales que $f(0) = f(1)$;
(c) todas las f tales que $f(3) = 1 + f(-5)$;
(d) todas las f tales que $f(-1) = 0$;
(e) todas las f que son continuas.

3. ¿Pertenece el vector $(3, -1, 0, -1)$ al subespacio de R^5 generado por los vectores $(2, -1, 3, 2)$, $(-1, 1, 1, -3)$ y $(1, 1, 9, -5)$?

4. Sea W el conjunto de todos los $(x_1, x_2, x_3, x_4, x_5)$ de R^5 que satisfacen

$$\begin{aligned} 2x_1 - x_2 + \tfrac{4}{3}x_3 - x_4 \qquad &= 0 \\ x_1 \qquad + \tfrac{2}{3}x_3 \qquad - x_5 &= 0 \\ 9x_1 - 3x_2 + 6x_3 - 3x_4 - 3x_5 &= 0. \end{aligned}$$

Encontrar un conjunto finito de vectores que genera W.

5. Sean F un cuerpo y n un entero positivo ($n \geq 2$). Sea V el espacio vectorial de todas las matrices $n \times n$ sobre F. ¿Cuáles de los siguientes conjuntos de matrices A de V son subespacios de V?

(a) todas las A inversibles;
(b) todas las A no inversibles;
(c) todas las A para las que $AB = BA$, donde B es cierta matriz dada de V;
(d) todas las A para las que $A^2 = A$.

6. (a) Demostrar que los únicos subespacios de R^1 son R^1 y el subespacio nulo.
(b) Demostrar que un subespacio de R^2 es R^2 o el subespacio nulo o consta de todos los múltiplos escalares de algún vector fijo de R^2. (El último tipo de subespacio es, intuitivamente, una recta por el origen.)
(c) ¿Puede usted describir los subespacios de R^3?

7. Sean W_1 y W_2 subespacios de un espacio vectorial V tal que la unión conjuntista de W_1 y W_2 sea también un subespacio. Demostrar que uno de los espacios W_i está contenido en el otro.

8. Sea V el espacio vectorial de todas las funciones de R en R; sea V_p el subconjunto de las funciones pares, $f(-x) = f(x)$; sea V_i el subconjunto de las funciones impares, $f(-x) = -f(x)$.

(a) Demostrar que V_p y V_i son subespacios de V.
(b) Demostrar que $V_p + V_i = V$.
(c) Demostrar que $V_p \cap V_i = \{0\}$.

9. Sea W_1 y W_2 subespacios de un espacio vectorial V tales que $W_1 + W_2 = V$ y $W_1 \cap W_2 = \{0\}$. Demostrar que para todo vector α de V existen únicos vectores α_1 en W_1 y α_2 en W_2 tales que $\alpha = \alpha_1 + \alpha_2$.

2.3. *Bases y dimensión*

Pasamos, ahora, a la tarea de dotar de dimensión a ciertos espacios vectoriales. Aunque usualmente asociamos «dimensión» con algo geométrico, debemos encontrar una definición algebraica apropiada para la dimensión de un espacio vectorial. Esto se hará mediante el concepto de base de un espacio vectorial.

Definición. *Sea V un espacio vectorial sobre F. Un subconjunto S de V se dice* **linealmente dependiente** (*o simplemente*, **dependiente**) *si existen vectores distintos $\alpha_1, \alpha_2, \ldots, \alpha_n$ de S y escalares $c_1, c_2, \ldots, c_n$ de F, no todos nulos, tales que*

$$c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n = 0.$$

Un conjunto que no es linealmente dependiente se dice **linealmente independiente.** *Si el conjunto S tiene solo un número finito de vectores $\alpha_1, \alpha_2, \ldots, \alpha_n$ se dice, a veces, que los $\alpha_1, \alpha_2, \ldots, \alpha_n$ son dependientes (o independientes), en vez de decir que S es dependiente (o independiente).*

Las siguientes son fáciles consecuencias de la definición.

1. Todo conjunto que contiene un conjunto linealmente dependiente es linealmente dependiente.
2. Todo subconjunto de un conjunto linealmente independiente es linealmente independiente.
3. Todo conjunto que contiene el vector 0 es linealmente dependiente; en efecto, $1 \cdot 0 = 0$.
4. Un conjunto S de vectores es linealmente independiente si, y solo si, todo subconjunto finito de S es linealmente independiente; es decir, si, y solo si para vectores diferentes $\alpha_1, \ldots, \alpha_n$ de S, arbitrariamente elegidos $c_1\alpha_1 + \cdots + c_n\alpha_n = 0$ implica que todo $c_i = 0$.

Definición. *Sea V un espacio vectorial. Una* **base** *de V es un conjunto de vectores linealmente independiente de V que genera el espacio V. El espacio V es de* **dimensión finita** *si tiene una base finita.*

Ejemplo 12. Sea F un subcuerpo de los números complejos. En F^3 los vectores

$$\begin{aligned} \alpha_1 &= (\ \ 3, 0, -3) \\ \alpha_2 &= (-1, 1, \ \ 2) \\ \alpha_3 &= (\ \ 4, 2, -2) \\ \alpha_4 &= (\ \ 2, 1, \ \ 1) \end{aligned}$$

son linealmente dependientes, pues

$$2\alpha_1 + 2\alpha_2 - \alpha_3 + 0 \cdot \alpha_4 = 0.$$

Los vectores

$$\begin{aligned} \epsilon_1 &= (1, 0, 0) \\ \epsilon_2 &= (0, 1, 0) \\ \epsilon_3 &= (0, 0, 1) \end{aligned}$$

son linealmente independientes.

Ejemplo 13. Sea F un cuerpo, y en F^n sea S el subconjunto que consta de los vectores $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ definidos por

$$\begin{aligned} \epsilon_1 &= (1, 0, 0, \ldots, 0) \\ \epsilon_2 &= (0, 1, 0, \ldots, 0) \\ &\cdots\cdots\cdots \\ \epsilon_n &= (0, 0, 0, \ldots, 1). \end{aligned}$$

Sean $x_1, x_2, \ldots, x_n$ escalares de F, y hágase $\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n$. Entonces

$$\alpha = (x_1, x_2, \ldots, x_n). \tag{2-12}$$

Esto muestra que $\epsilon_1, \ldots, \epsilon_n$ generan F^n. Como $\alpha = 0$ si, y solo si, $x_1 = x_2 = \ldots = x_n = 0$, los vectores $\epsilon_1, \ldots, \epsilon_n$ son linealmente independientes. El conjunto

$S = \{\epsilon_1, \ldots, \epsilon_n\}$ es, por tanto, una base de F^n. Esta base particular se llamará **base canónica** de F^n.

Ejemplo 14. Sea P una matriz $n \times n$ inversible con elementos en el cuerpo F. Entonces $P_1, \ldots, P_n$, las columnas de P, forman una base del espacio de las matrices columnas $F^{n \times 1}$. Eso se verá como sigue. Si X es una matriz columna, entonces

$$PX = x_1P_1 + \cdots + x_nP_n.$$

Como $PX = 0$ tiene solo la solución trivial $X = 0$, se sigue que $\{P_1, \ldots, P_n\}$ es un conjunto linealmente independiente. ¿Por qué generan $F^{n \times 1}$? Sea Y cualquier matriz columna. Si $X = P^{-1}Y$, entonces $Y = PX$, esto es

$$Y = x_1P_1 + \cdots + x_nP_n.$$

Así, $\{P_1, \ldots, P_n\}$ es una base de $F^{n \times 1}$.

Ejemplo 15. Sea A una matriz $m \times n$ y sea S el espacio solución del sistema homogéneo $AX = 0$ (Ejemplo 7). Sea R una matriz escalón reducida por filas que es equivalente por filas a A. Entonces S es también el espacio solución del sistema $RX = 0$. Si R tiene r filas no nulas, entonces el sistema de ecuaciones $RX = 0$ simplemente expresa r de las incógnitas $x_1, \ldots, x_n$ en términos de las $(n - r)$ incógnitas x_j restantes. Supóngase que los elementos principales, no nulos, de las filas no nulas están en las columnas $k_1, \ldots, k_r$. Sea J el conjunto constituido por los $n - r$ índices diferentes de $k_1, \ldots, k_r$:

$$J = \{1, \ldots, n\} - \{k_1, \ldots, k_r\}.$$

El sistema $RX = 0$ tiene la forma

$$\begin{array}{ccc} x_{k_1} + \sum_J c_{1j}x_j = 0 \\ \vdots \quad \vdots \quad \vdots \\ x_{k_r} + \sum_J c_{rj}x_j = 0 \end{array}$$

donde los c_{ij} son ciertos escalares. Todas las soluciones se obtienen asignando (arbitrariamente) valores a aquellos x_j con j en J y calculando los correspondientes valores de $x_{k1}, \ldots, x_{kr}$. Para cada j de J, sea E_j la solución obtenida al hacer $x_j = 1$ y $x_i = 0$, para todos los otros i de J. Se afirma que los $(n - r)$ vectores E_j, con j en J, forman una base para el espacio solución.

Como la matriz columna E_j tiene un 1 en la fila j y ceros en las restantes filas, las razones dadas en el Ejemplo 13 permiten concluir que el conjunto de estos vectores es linealmente independiente. Por esta razón ese espacio genera el espacio solución. Si la matriz columna T, con elementos $t_1, \ldots, t_n$, pertenece al espacio solución, la matriz

$$N = \sum_J t_jE_j$$

pertenece también al espacio solución, y es una solución tal que $x_j = t_j$ para cada j de J. La solución con esta propiedad es única; luego $N = T$ y T pertenece al espacio generado por los vectores E_j.

Ejemplo 16. Daremos ahora un ejemplo de una base infinita. Sea F un subcuerpo de los números complejos y sea V el espacio de las funciones polinomios sobre F. Se recuerda que estas funciones son las de F en F que tienen la forma

$$f(x) = c_0 + c_1x + \cdots + c_nx^n.$$

Sea $f_k(x) = x^k$, $k = 0, 1, 2, \ldots$ El conjunto (infinito) $\{f_0, f_1, f_2, \ldots\}$ es una base de V. Es claro que el conjunto genera V, pues la función f anterior es

$$f = c_0f_0 + c_1f_1 + \cdots + c_nf_n.$$

El lector verá que esto es virtualmente una repetición de la definición de la función polinomio, es decir, una función de F en F es una función polinomio si, y solo si, existe un entero n y escalares $c_0, \ldots, c_n$ tales que $f = c_0f_0 + \cdots + c_nf_n$. ¿Por qué son estas funciones independientes? Demostrar que el conjunto $\{f_0, f_1, f_2, \ldots\}$ es independiente, es lo mismo que demostrar que todo subconjunto finito suyo es independiente. Bastará entonces demostrar que, para todo n, el conjunto $\{f_0, \ldots, f_n\}$ es independiente. Supóngase que

$$c_0f_0 + \cdots + c_nf_n = 0.$$

Esto dice que

$$c_0 + c_1x + \cdots + c_nx^n = 0$$

para todo x de F; en otras palabras, cada x de F es raíz del polinomio $f(x) = c_0 + c_1x + \cdots + c_nx^n$. Se supone que el lector sabe que un polinomio de grado n con coeficientes complejos puede tener a lo sumo n raíces distintas. Se sigue que, $c_0 = c_1 = \cdots = c_n = 0$.

Se ha visto una base infinita de V. ¿Quiere decir esto que V no es de dimensión finita? En realidad, es así, pero ello no se deduce de la definición, porque de todo lo que se sabe V podría tener también una base finita. Esa posibilidad queda fácilmente descartada. (Se eliminará en general en el siguiente teorema.) Supóngase que tenemos un número finito de funciones polinomios $g_1, \ldots, g_r$. Habrá una mayor potencia de x que aparece (con coeficiente no nulo) en $g_1(x), \ldots, g_r(x)$. Si tal potencia es k, es claro que $f_{k+1}(x) = x^{k+1}$ no pertenece al subespacio generado por los $g_1, \ldots, g_r$. Así, pues, V no es de dimensión finita.

Un comentario final respecto a este ejemplo. Una base infinita no tiene nada que ver con «combinaciones lineales infinitas». El lector que sienta el deseo irresistible de introducir series de potencias

$$\sum_{k=0}^{\infty} c_kx^k$$

en este ejemplo, debe estudiarlo cuidadosamente de nuevo. Si con ello no se le aclara todavía, debe limitar su atención, de ahora en adelante, a espacios de dimensión finita.

Teorema 4. *Sea V un espacio vectorial generado por un conjunto finito de vectores $\beta_1, \beta_2, \ldots, \beta_m$. Entonces todo conjunto independiente de vectores de V es finito y no contiene más de m elementos.*

Demostración. Para demostrar este teorema es suficiente ver que todo subconjunto S de V que contiene más de m vectores es linealmente dependiente. Sea S un tal conjunto. En S hay vectores diferentes $\alpha_1, \alpha_2, \ldots, \alpha_n$, donde $n > m$. Como $\beta_1, \ldots, \beta_m$ generan V, existen escalares A_{ij} en F tales que

$$\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i.$$

Para cualesquiera n escalares $x_1, x_2, \ldots, x_n$, se tiene

$$\begin{aligned} x_1\alpha_1 + \cdots + x_n\alpha_n &= \sum_{j=1}^{n} x_j\alpha_j \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\beta_i \\ &= \sum_{j=1}^{n} \sum_{i=1}^{m} (A_{ij}x_j)\beta_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j\right)\beta_i. \end{aligned}$$

Como $n > m$, el Teorema 6 del Capítulo 1 implica que existen escalares $x_1, x_2, \ldots, x_n$, no todos 0, tales que

$$\sum_{j=1}^{n} A_{ij}x_j = 0, \qquad 1 \leq i \leq m.$$

Luego $x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n = 0$. Ello demuestra que S es un conjunto linealmente dependiente. ▮

Corolario 1. *Si V es un espacio vectorial de dimensión finita, entonces dos bases cualesquiera de V tienen el mismo número (finito) de elementos.*

Demostración. Como V es de dimensión finita, tiene una base finita

$$\{\beta_1, \beta_2, \ldots, \beta_m\}.$$

Por el Teorema 4, toda base de V es finita y contiene no más de m elementos. Así, si $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ es una base, $n \leq m$. Por el mismo razonamiento, $m \leq n$. Luego $m = n$. ▮

Este corolario permite definir la **dimensión** de un espacio vectorial de dimensión finita como el número de elementos de una base cualquiera de V. Se indicará la dimensión de un espacio V de dimensión finita por dim V. Ello nos permite volver a enunciar el Teorema 4 como sigue.

Corolario 2. *Sea V un espacio vectorial de dimensión finita y sea $n = \dim V$. Entonces*

(a) *cualquier subconjunto de V que contenga más de n vectores es linealmente dependiente;*

(b) *ningún subconjunto de V que contenga menos de n vectores puede generar V.*

Ejemplo 17. Si F es un cuerpo, la dimensión de F^n es n, porque la base canónica de F^n contiene n vectores. El espacio de las matrices $F^{m \times n}$ tiene dimensión mn. Ello es claro por analogía con el caso de F^n, ya que las mn matrices que tienen un 1 en el lugar i, j, con ceros en los demás, forman una base de $F^{m \times n}$. Si A es una matriz $m \times n$, entonces el espacio solución de A tiene dimensión $n - r$, donde r es el número de filas no nulas de una matriz escalón reducida por filas, que es equivalente por filas a A. Véase el Ejemplo 15.

Si V es cualquier espacio vectorial sobre F, el subespacio nulo de V es generado por el vector 0, pero $\{0\}$ es un conjunto linealmente dependiente y no una base. Por esta razón se conviene que el subespacio nulo tenga dimensión 0. Se podría llegar también a la misma conclusión pensando que el conjunto vacío es una base del subespacio nulo. El conjunto vacío genera $\{0\}$, pues la intersección de todos los subespacios que contiene el conjunto vacío es $\{0\}$, y el conjunto vacío es linealmente independiente, porque no contiene vectores.

Lema. *Sea S un subconjunto linealmente independiente de un espacio vectorial V. Supóngase que β es un vector de V que no pertenece al subespacio generado por S. Entonces el conjunto que se obtiene agregando β a S, es linealmente independiente.*

Demostración. Supóngase que $\alpha_1, \ldots, \alpha_m$ son vectores distintos de S y que

$$c_1\alpha_1 + \cdots + c_m\alpha_m + b\beta = 0.$$

Entonces $b = 0$; de otra manera,

$$\beta = \left(-\frac{c_1}{b}\right)\alpha_1 + \cdots + \left(-\frac{c_m}{b}\right)\alpha_m$$

y β está en el subespacio generado por S. Así, $c_1\alpha_1 + \cdots + c_m\alpha_m = 0$, y como S es un conjunto linealmente independiente, todo $c_i = 0$. ∎

Teorema 5. *Si W es un subespacio de un espacio vectorial de dimensión finita V, todo subconjunto linealmente independiente de W es finito y es parte de una base (finita) de W.*

Demostración. Supóngase que S_0 es un conjunto linealmente independiente de W. Si S es un subconjunto linealmente independiente de W que contiene a S_0, entonces S es también un subconjunto linealmente independiente de V; como V es de dimensión finita, S no tiene más de dim V elementos.

Se extiende S_0 a una base de W, como sigue. Si S_0 genera W, entonces S_0 es una base de W y está demostrado. Si S_0 no genera W, por el lema anterior se halla un vector β_1 en W tal que el conjunto $S_1 = S_0 \cup \{\beta_1\}$ es independiente. Si S_1 genera W, está demostrado. Si no, se aplica el lema para obtener un vector β_2

en W tal que $S_2 = S_1 \cup \{\beta_2\}$ es independiente. Si se continúa de este modo, entonces (y en no más de dim V de etapas) se llega a un conjunto

$$S_m = S_0 \cup \{\beta_1, \ldots, \beta_m\}$$

que es una base de W. ∎

Corolario 1. *Si W es un subespacio propio de un espacio vectorial de dimensión finita V, entonces W es de dimensión finita y dim $W <$ dim V.*

Demostración. Podemos suponer que W contiene un vector $\alpha \neq 0$. Por el Teorema 5 y su demostración existe una base de W que, conteniendo a α, no contiene más que dim V elementos. Luego W es de dimensión finita y dim $W \leq$ dim V. Como es un subespacio propio existe un vector β en V que no está en W. Agregando β a cualquier base de W se obtiene un subconjunto linealmente independiente de V. Así, dim $W <$ dim V. ∎

Corolario 2. *En un espacio vectorial V de dimensión finita todo conjunto linealmente independiente de vectores es parte de una base.*

Corolario 3. *Sea A una matriz $n \times n$ sobre el cuerpo F, y supóngase que los vectores fila de A forman un conjunto linealmente independiente de vectores de F^n. Entonces A es inversible.*

Demostración. Sean $\alpha_1, \alpha_2, \ldots, \alpha_n$ vectores fila de A, y supóngase que W es un subespacio de F^n generado por $\alpha_1, \alpha_2, \ldots, \alpha_n$. Como los $\alpha_1, \alpha_2, \ldots, \alpha_n$ son linealmente independientes, la dimensión de W es n. Por el Corolario 1 se tiene que $W = F^n$. Luego existen escalares B_{ij} en F tales que

$$\epsilon_i = \sum_{j=1}^{n} B_{ij}\alpha_j, \qquad 1 \leq i \leq n$$

donde $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ es la base canónica de F^n. Así, para la matriz B de elementos B_{ij} se tiene

$$BA = I. \quad ∎$$

Teorema 6. *Si W_1 y W_2 son subespacios de dimensión finita de un espacio vectorial, entonces $W_1 + W_2$ es de dimensión finita y*

$$\dim W_1 + \dim W_2 = \dim (W_1 \cap W_2) + \dim (W_1 + W_2).$$

Demostración. Por el Teorema 5 y sus corolarios, $W_1 \cap W_2$ tiene una base finita $\{\alpha_1, \ldots, \alpha_k\}$ que es parte de la base

$$\{\alpha_1, \ldots, \alpha_k, \ \beta_1, \ldots, \beta_m\} \qquad \text{para } W_1$$

y parte de la base

$$\{\alpha_1, \ldots, \alpha_k, \ \gamma_1, \ldots, \gamma_n\} \qquad \text{para } W_2.$$

El subespacio $W_1 + W_2$ es generado por los vectores

$$\alpha_1, \ldots, \alpha_k, \qquad \beta_1, \ldots, \beta_m, \qquad \gamma_1, \ldots, \gamma_n$$

y estos vectores forman un conjunto independiente. En efecto, supóngase que

$$\Sigma x_i\alpha_i + \Sigma y_j\beta_j + \Sigma z_r\gamma_r = 0.$$

Entonces

$$-\Sigma z_r\gamma_r = \Sigma x_i\alpha_i + \Sigma y_j\beta_j$$

que muestra que $\Sigma z_r\gamma_r$ pertenece a W_1. Como $\Sigma z_r\gamma_r$ también pertenece a W_2, se sigue que

$$\Sigma z_r\gamma_r = \Sigma c_i\alpha_i$$

para ciertos escalares $c_1, \ldots, c_k$. Como el conjunto

$$\{\alpha_1, \ldots, \alpha_k, \quad \gamma_1, \ldots, \gamma_n\}$$

es independiente, cada uno de los escalares $z_r = 0$. Así,

$$\Sigma x_i\alpha_i + \Sigma y_j\beta_j = 0$$

y como

$$\{\alpha_1, \ldots, \alpha_k, \quad \beta_1, \ldots, \beta_m\}$$

es también un conjunto independiente, cada $x_i = 0$ y cada $y_i = 0$. Así,

$$\{\alpha_1, \ldots, \alpha_k, \quad \beta_1, \ldots, \beta_m, \quad \gamma_1, \ldots, \gamma_n\}$$

es una base para $W_1 + W_2$. Finalmente,

$$\begin{aligned} \dim W_1 + \dim W_2 &= (k + m) + (k + n) \\ &= k + (m + k + n) \\ &= \dim (W_1 \cap W_2) + \dim (W_1 + W_2). \quad \blacksquare \end{aligned}$$

Cerremos esta sección con una observación referente a la independencia y dependencia lineal. Se han definido estos conceptos para conjuntos de vectores. Es útil haberlos definido para sucesiones finitas (n-tuples ordenados) de vectores: $\alpha_1, \ldots, \alpha_n$. Se dirá que los vectores $\alpha_1, \ldots, \alpha_n$ son **linealmente dependientes** si existen escalares fijos $c_1, \ldots, c_n$, no todos nulos, tales que $c_1\alpha_1 + \cdots + c_n\alpha_n = 0$. Todo esto es tan natural que el lector podría creer que se ha usado ya esta terminología. ¿Cuál es la diferencia entre una sucesión finita $\alpha_1, \ldots, \alpha_n$ y el conjunto $\{\alpha_1, \ldots, \alpha_n\}$? Hay dos diferencias: identidad y orden.

Si se examina el conjunto $\{\alpha_1, \ldots, \alpha_n\}$ es corriente suponer que no hay dos vectores $\alpha_1, \ldots, \alpha_n$ que sean idénticos. En una sucesión $\alpha_1, \ldots, \alpha_n$ todos los α_i pueden ser el mismo vector. Si $\alpha_i = \alpha_j$, para algún $i \neq j$, entonces la sucesión $\alpha_1, \ldots, \alpha_n$ es linealmente dependiente:

$$\alpha_i + (-1)\alpha_j = 0.$$

Así, si $\alpha_1, \ldots, \alpha_n$ son linealmente independientes, todos son distintos, y se puede hablar del conjunto $\{\alpha_1, \ldots, \alpha_n\}$ sabiendo que tiene n vectores. Así, evidentemente, ninguna confusión surgirá en el estudio de bases y dimensión. La dimensión de un espacio V de dimensión finita es el mayor n, tal que un

n-tuple de vectores de V es linealmente independiente, y así sucesivamente. El lector que crea que este párrafo no es más que palabras, puede preguntarse si los vectores

$$\alpha_1 = (e^{\pi/2}, 1)$$
$$\alpha_2 = (\sqrt[3]{110}, 1)$$

son linealmente independientes en R^2.

Los elementos de una sucesión están enumerados en un orden determinado. Un conjunto es una colección de objetos con una disposición no determinada u ordenada. Claro que al describir el conjunto se deben indicar sus elementos, y ello requiere elegir un orden. Pero el orden no es parte del conjunto. Los conjuntos $\{1, 2, 3, 4\}$ y $\{4, 3, 2, 1\}$ son idénticos, mientras que 1, 2, 3, 4 es una sucesión muy diferente de 4, 3, 2, 1. El aspecto ordinal de las sucesiones no entra en juego en los asuntos de independencia, dependencia, etc., porque la dependencia (como se definió) no está afectada por el orden. La sucesión $\alpha_n, \ldots, \alpha_1$ es dependiente si, y solo si, la sucesión $\alpha_1, \ldots, \alpha_n$ es dependiente. En la sección siguiente, el orden será importante.

Ejercicios

1. Demostrar que si dos vectores son linealmente dependientes, uno de ellos es un múltiplo escalar del otro.

2. ¿Son los vectores

$$\alpha_1 = (1, 1, 2, 4), \qquad \alpha_2 = (2, -1, -5, 2)$$
$$\alpha_3 = (1, -1, -4, 0), \qquad \alpha_4 = (2, 1, 1, 6)$$

linealmente independientes en R^4?

3. Hallar una base para el subespacio de R^4 generado por los cuatro vectores del Ejercicio 2.

4. Demostrar que los vectores

$$\alpha_1 = (1, 0, -1), \qquad \alpha_2 = (1, 2, 1), \qquad \alpha_3 = (0, -3, 2)$$

forman una base para R^3. Expresar cada uno de los vectores de la base canónica como combinación lineal de α_1, α_2 y α_3.

5. Hallar tres vectores de R^3 que sean linealmente dependientes y tales que dos cualesquiera de ellos sean linealmente independientes.

6. Sea V el espacio vectorial de todas las matrices 2×2 sobre el cuerpo F. Demostrar que V tiene dimensión 4, encontrando una base de V que tenga cuatro elementos.

7. Sea V el espacio vectorial del Ejercicio 6. Sea W_1 el conjunto de las matrices de la forma

$$\begin{bmatrix} x & -x \\ y & z \end{bmatrix}$$

y sea W_2 el conjunto de las matrices de la forma

$$\begin{bmatrix} a & b \\ -a & c \end{bmatrix}.$$

(a) Demostrar que W_1 y W_2 son subespacios de V.
(b) Hallar la dimensión de W_1, W_2, $W_1 + W_2$ y $W_1 \cap W_2$.

8. Nuevamente, sea V el espacio de las matrices 2×2 sobre F. Hallar una base $\{A_1, A_2, A_3, A_4\}$ de V, de modo que $A_j^2 = A_j$ para cada j.

9. Sea V un espacio vectorial sobre el cuerpo F de los números complejos. Supóngase que α, β y γ sean vectores linealmente independientes en V. Demostrar que $(\alpha + \beta)$, $(\beta + \gamma)$ y $(\gamma + \alpha)$ son linealmente independientes.

10. Sea V un espacio vectorial sobre el cuerpo F. Supóngase que hay un número finito de vectores $\alpha_1, \ldots, \alpha_r$ en V que generan V. Demostrar que V es de dimensión finita.

11. Sea V el conjunto de todas las matrices 2×2, A, con elementos *complejos* que satisfacen $A_{11} + A_{22} = 0$.
(a) Hacer ver que V es un espacio vectorial sobre el cuerpo de los números reales, con las operaciones comunes de adición de matrices y multiplicación de matrices por un escalar.
(b) Hallar una base de este espacio vectorial.
(c) Sea W el conjunto de todas las matrices A de V tales que $A_{21} = -\bar{A}_{21}$ (la barra indica complejo conjugado). Demostrar que W es un subespacio de V y hallar una base de W.

12. Demostrar que el espacio de todas las matrices $m \times n$ sobre el cuerpo F tiene dimensión mn, presentando una base para este espacio.
13. Analizar el Ejercicio 9, cuando V es un espacio vectorial sobre el cuerpo de dos elementos, descrito en el Ejercicio 5, Sección 1.1.

14. Sea V el conjunto de los números reales. Considerar V como un espacio vectorial sobre el cuerpo de los números *racionales*, con las operaciones usuales. Demostrar que este espacio vectorial *no* es de dimensión finita.

2.4. *Coordenadas*

Una de las características útiles de una base $\mathcal{B}$ en un espacio V de dimensión n es que permite esencialmente introducir coordenadas en V en forma análoga a las «coordenadas naturales», x_i, de un vector $\alpha = (x_1, \ldots, x_n)$ en el espacio F^n. En este esquema, las coordenadas de un vector α en V, respecto de la base $\mathcal{B}$, serán los escalares que sirven para expresar α como combinación lineal de los vectores de la base. Así, pues, sería de considerar las coordenadas naturales de un vector α en F^n como definidas por α y la base canónica de F^n; sin embargo, al adoptar este punto de vista se debe tener mucho cuidado. Si

$$\alpha = (x_1, \ldots, x_n) = \Sigma\, x_i \epsilon_i$$

y $\mathcal{B}$ es la base canónica de F^n, ¿cómo quedan determinadas las coordenadas de α por $\mathcal{B}$ y α? Una manera de formular la respuesta es la siguiente. Un vector dado, α, tiene una expresión única como combinación lineal de los vectores de la base canónica, y la coordenada i-ésima x_i de α es el coeficiente de ϵ_i en

esta expresión. Desde este punto de vista, se puede decir cuál es la i-ésima coordenada, porque se tiene un orden «natural» de los vectores de la base canónica, esto es, se tiene una regla para determinar cuál es el «primer» vector en la base, cuál es el «segundo» vector, y así sucesivamente. Si $\mathcal{B}$ es una base arbitraria del espacio V de dimensión n, no se tendrá probablemente, un orden natural de los vectores de $\mathcal{B}$, y, por tanto, será necesario imponer algún orden a estos vectores antes que se pueda definir la «i-ésima coordenada de α respecto a $\mathcal{B}$». Para plantearlo de otra forma, se definirán las coordenadas respecto a una sucesión de vectores y no de un conjunto de vectores.

Definición. *Si V es un espacio vectorial de dimensión finita, una* **base ordenada** *de V es una sucesión finita de vectores linealmente independiente y que genera V.*

Si la sucesión $\alpha_1, \ldots, \alpha_n$ es una base ordenada de V, entonces el conjunto $\{\alpha_1, \ldots, \alpha_n\}$ es una base de V. La base ordenada es el conjunto, juntamente con el orden dado. Se incurrirá en un pequeño abuso de notación y se escribirá

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$$

diciendo que $\mathcal{B}$ es una base ordenada de V.

Ahora supóngase que V es un espacio vectorial de dimensión finita sobre el cuerpo F y que

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}.$$

es una base ordenada de V. Dado α de V, existe un único n-tuple $(x_1, \ldots, x_n)$ de escalares tal que

$$\alpha = \sum_{i=1}^{n} x_i\alpha_i.$$

Este n-tuple es único, ya que si también se tiene

$$\alpha = \sum_{i=1}^{n} z_i\alpha_i$$

entonces

$$\sum_{i=1}^{n} (x_i - z_i)\alpha_i = 0$$

y la independencia lineal de los α_i asegura que $x_i - z_i = 0$, para todo i. Se llamará a x_i la i-ésima **coordenada de α respecto a la base ordenada**

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$$

Si

$$\beta = \sum_{i=1}^{n} y_i\alpha_i$$

entonces

$$\alpha + \beta = \sum_{i=1}^{n} (x_i + y_i)\alpha_i$$

de modo que la i-ésima coordenada de $(\alpha + \beta)$ en esta base ordenada es $(x_i + y_i)$. En forma análoga, la i-ésima coordenada de $(c\alpha)$ es cx_i. Debe observarse también que, cada n-tuple $(x_1, \ldots, x_n)$ de F^n es el n-tuple de coordenadas de algún vector de V, a saber, el vector

$$\sum_{i=1}^{n} x_i\alpha_i.$$

Resumiendo, cada base ordenada de V determina una correspondencia biunívoca

$$\alpha \rightarrow (x_1, \ldots, x_n)$$

entre el conjunto de todos los vectores de V y el conjunto de todos los n-tuples de F^n. Esta correspondencia tiene la propiedad de que el correspondiente de $(\alpha + \beta)$ es la suma en F^n de los correspondientes de α y β, y el correspondiente de $(c\alpha)$ es el producto en F^n del escalar c y el correspondiente de α.

Cabría preguntarse, en este momento, ¿por qué no se elige simplemente alguna base ordenada de V y se expresa todo V por su correspondiente n-tuple de coordenadas, ya que así se tiene la conveniencia de operar solo con n-tuples? Esto iría contra nuestro propósito por dos razones. Primera, como la definición axiomática de espacios vectoriales indica, se trata de aprender a razonar con espacios vectoriales como sistemas algebraicos abstractos. Segunda, incluso en aquellos casos en que se usan coordenadas, los resultados más importantes se obtienen de la capacidad que se tenga de cambiar sistemas de coordenadas, es decir, de cambiar la base ordenada.

A menudo, será de mayor conveniencia usar la **matriz de las coordenadas de α respecto a la base ordenada $\mathcal{B}$**:

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

en vez del n-tuple $(x_1, \ldots, x_n)$ de coordenadas. Para indicar la dependencia de esta matriz de coordenadas respecto de la base se usará el símbolo

$$[\alpha]_{\mathcal{B}}$$

para la matriz de coordenadas del vector α respecto a la base ordenada $\mathcal{B}$. Esta notación será particularmente útil cuando procedamos ahora a describir qué le sucede a las coordenadas de un vector α cuando se cambia de una base ordenada a otra.

Supóngase, entonces, que V es de dimensión n y que

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{y} \quad \mathcal{B}' = \{\alpha_1', \ldots, \alpha_n'\}$$

sean dos bases ordenadas de V. Existen escalares únicos P_{ij} tales que

$$\text{(2-13)} \qquad \alpha_j' = \sum_{i=1}^{n} P_{ij}\alpha_i, \qquad 1 \leq j \leq n.$$

Sean $x_1', \ldots, x_n'$ las coordenadas de un vector dado α en la base ordenada $\mathcal{B}'$. Entonces

$$\begin{aligned}\alpha &= x_1'\alpha_1' + \cdots + x_n'\alpha_n' \\ &= \sum_{j=1}^{n} x_j'\alpha_j' \\ &= \sum_{j=1}^{n} x_j' \sum_{i=1}^{n} P_{ij}\alpha_i \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} (P_{ij}x_j')\alpha_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} P_{ij}x_j'\right) \alpha_i.\end{aligned}$$

Se tiene así la relación

(2-14) $$\alpha = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} P_{ij}x_j'\right) \alpha_i.$$

Como las coordenadas $x_1, x_2, \ldots, x_n$ de α en la base ordenada $\mathcal{B}$ están unívocamente determinadas, se sigue de (2-14) que

(2-15) $$x_i = \sum_{j=1}^{n} P_{ij}x_j', \qquad 1 \leq i \leq n.$$

Sea P la matriz $n \times n$ cuyo i, j elemento es el escalar P_{ij} y sean X y X' las matrices de coordenadas del vector α en las bases ordenadas $\mathcal{B}$ y $\mathcal{B}'$. Entonces se puede escribir (2-15) como

(2-16) $$X = PX'.$$

Como $\mathcal{B}$ y $\mathcal{B}'$ son conjuntos linealmente independientes, $X = 0$, si, y solo si, $X' = 0$. Así, de (2-16) y del Teorema 7 del Capítulo 1, se sigue que P es inversible. Luego

(2-17) $$X' = P^{-1}X.$$

Si se usa la notación introducida anteriormente para la matriz de coordenadas de un vector respecto a una base ordenada, entonces (2-16) y (2-17) dicen que

$$\begin{aligned}[\alpha]_{\mathcal{B}} &= P[\alpha]_{\mathcal{B}'} \\ [\alpha]_{\mathcal{B}'} &= P^{-1}[\alpha]_{\mathcal{B}}.\end{aligned}$$

Con lo que lo anterior puede resumirse así.

Teorema 7. *Sea V un espacio vectorial de dimensión n sobre el cuerpo F, y sean $\mathcal{B}$ y $\mathcal{B}'$ bases ordenadas de V. Entonces existe una única matriz $n \times n$, necesariamente inversible, con elementos de F, de modo que*

(i) $$[\alpha]_{\mathcal{B}} = P[\alpha]_{\mathcal{B}'}$$
(ii) $$[\alpha]_{\mathcal{B}'} = P^{-1}[\alpha]_{\mathcal{B}}$$

para todo vector α de V. Las columnas de P están dadas por

$$P_j = [\alpha_j']_{\mathcal{B}}, \qquad j = 1, \ldots, n.$$

Para completar el análisis anterior, se demostrará también el siguiente teorema.

Teorema 8. *Supóngase que P es una matriz inversible $n \times n$ sobre F. Sea V un espacio vectorial de dimensión n sobre F, y sea $\mathcal{B}$ una base ordenada de V. Entonces existe una base ordenada $\mathcal{B}'$, única, de V tal que*

(i) $$[\alpha]_{\mathcal{B}} = P[\alpha]_{\mathcal{B}'}$$
(ii) $$[\alpha]_{\mathcal{B}'} = P^{-1}[\alpha]_{\mathcal{B}}$$

para todo vector α en V.

Demostración. Sea $\mathcal{B}$, que consta de los vectores $\alpha_1, \ldots, \alpha_n$. Si $\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ es una base ordenada de V, para la cual se tiene (i), es claro que

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Así, solo se necesita demostrar que los vectores α'_j, definidos por estas igualdades, forman una base. Sea $Q = P^{-1}$. Entonces

$$\begin{aligned}\sum_j Q_{jk}\alpha'_j &= \sum_j Q_{jk} \sum_i P_{ij}\alpha_i \\ &= \sum_j \sum_i P_{ij}Q_{jk}\,\alpha_i \\ &= \sum_i \Big(\sum_j P_{ij}Q_{jk}\Big)\alpha_i \\ &= \alpha_k.\end{aligned}$$

Con lo que el subespacio generado por el conjunto

$$\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

contiene a $\mathcal{B}$ y, por tanto, es igual a V. Así que $\mathcal{B}'$ es una base, y por su definición y por el Teorema 7 es claro que (i) es válido y también lo es (ii). ∎

Ejemplo 18. Sea F un cuerpo y sea

$$\alpha = (x_1, x_2, \ldots, x_n)$$

un vector de F^n. Si $\mathcal{B}$ es la base ordenada canónica de F^n,

$$\mathcal{B} = \{\epsilon_1, \ldots, \epsilon_n\}$$

la matriz de coordenadas del vector α en la base $\mathcal{B}$ está dada por

$$[\alpha]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Ejemplo 19. Sea R el cuerpo de los números reales y sea θ un número real dado. La matriz

$$P = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$$

$$\begin{aligned}\alpha &= x'_1\alpha'_1 + \cdots + x'_n\alpha'_n \\ &= \sum_{j=1}^{n} x'_j\alpha'_j \\ &= \sum_{j=1}^{n} x'_j \sum_{i=1}^{n} P_{ij}\alpha_i \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} (P_{ij}x'_j)\alpha_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} P_{ij}x'_j\right) \alpha_i.\end{aligned}$$

Se tiene así la relación

(2-14) $$\alpha = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} P_{ij}x'_j\right) \alpha_i.$$

Como las coordenadas $x_1, x_2, \ldots, x_n$ de α en la base ordenada $\mathcal{B}$ están unívocamente determinadas, se sigue de (2-14) que

(2-15) $$x_i = \sum_{j=1}^{n} P_{ij}x'_j, \qquad 1 \leq i \leq n.$$

Sea P la matriz $n \times n$ cuyo i, j elemento es el escalar P_{ij} y sean X y X' las matrices de coordenadas del vector α en las bases ordenadas $\mathcal{B}$ y $\mathcal{B}'$. Entonces se puede escribir (2-15) como

(2-16) $$X = PX'.$$

Como $\mathcal{B}$ y $\mathcal{B}'$ son conjuntos linealmente independientes, $X = 0$, si, y solo si, $X' = 0$. Así, de (2-16) y del Teorema 7 del Capítulo 1, se sigue que P es inversible. Luego

(2-17) $$X' = P^{-1}X.$$

Si se usa la notación introducida anteriormente para la matriz de coordenadas de un vector respecto a una base ordenada, entonces (2-16) y (2-17) dicen que

$$\begin{aligned}[][\alpha]_{\mathcal{B}} &= P[\alpha]_{\mathcal{B}'} \\ [\alpha]_{\mathcal{B}'} &= P^{-1}[\alpha]_{\mathcal{B}}.\end{aligned}$$

Con lo que lo anterior puede resumirse así.

Teorema 7. *Sea V un espacio vectorial de dimensión n sobre el cuerpo F, y sean $\mathcal{B}$ y $\mathcal{B}'$ bases ordenadas de V. Entonces existe una única matriz $n \times n$, necesariamente inversible, con elementos de F, de modo que*

(i) $$[\alpha]_{\mathcal{B}} = P[\alpha]_{\mathcal{B}'}$$
(ii) $$[\alpha]_{\mathcal{B}'} = P^{-1}[\alpha]_{\mathcal{B}}$$

para todo vector α de V. Las columnas de P están dadas por

$$P_j = [\alpha'_j]_{\mathcal{B}}, \qquad j = 1, \ldots, n.$$

Para completar el análisis anterior, se demostrará también el siguiente teorema.

Teorema 8. *Supóngase que P es una matriz inversible n × n sobre F. Sea V un espacio vectorial de dimensión n sobre F, y sea* $\mathcal{B}$ *una base ordenada de V. Entonces existe una base ordenada* $\mathcal{B}'$*, única, de V tal que*

(i) $$[\alpha]_{\mathcal{B}} = P[\alpha]_{\mathcal{B}'}$$
(ii) $$[\alpha]_{\mathcal{B}'} = P^{-1}[\alpha]_{\mathcal{B}}$$

para todo vector α *en V.*

Demostración. Sea $\mathcal{B}$, que consta de los vectores $\alpha_1, \ldots, \alpha_n$. Si $\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ es una base ordenada de V, para la cual se tiene (i), es claro que

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Así, solo se necesita demostrar que los vectores α'_j, definidos por estas igualdades, forman una base. Sea $Q = P^{-1}$. Entonces

$$\begin{aligned}\sum_j Q_{jk}\alpha'_j &= \sum_j Q_{jk} \sum_i P_{ij}\alpha_i \\ &= \sum_j \sum_i P_{ij}Q_{jk}\,\alpha_i \\ &= \sum_i \Big(\sum_j P_{ij}Q_{jk}\Big)\alpha_i \\ &= \alpha_k.\end{aligned}$$

Con lo que el subespacio generado por el conjunto

$$\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

contiene a $\mathcal{B}$ y, por tanto, es igual a V. Así que $\mathcal{B}'$ es una base, y por su definición y por el Teorema 7 es claro que (i) es válido y también lo es (ii). ▌

Ejemplo 18. Sea F un cuerpo y sea

$$\alpha = (x_1, x_2, \ldots, x_n)$$

un vector de F^n. Si $\mathcal{B}$ es la base ordenada canónica de F^n,

$$\mathcal{B} = \{\epsilon_1, \ldots, \epsilon_n\}$$

la matriz de coordenadas del vector α en la base $\mathcal{B}$ está dada por

$$[\alpha]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Ejemplo 19. Sea R el cuerpo de los números reales y sea θ un número real dado. La matriz

$$P = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$$

es inversible con inversa,

$$P^{-1} = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

Así que para cada θ, el conjunto $\mathcal{B}'$ que consta de los vectores $(\cos\theta, \operatorname{sen}\theta)$, $(\operatorname{sen}\theta, \cos\theta)$ es una base de R^2; intuitivamente esta base puede ser descrita como la obtenida por rotación de ángulo θ de la base canónica. Si α es el vector (x_1, x_2), entonces

$$[\alpha]_{\mathcal{B}'} = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

o

$$\begin{aligned} x_1' &= x_1 \cos\theta + x_2 \operatorname{sen}\theta \\ x_2' &= -x_1 \operatorname{sen}\theta + x_2 \cos\theta. \end{aligned}$$

Ejemplo 20. Sea F un subcuerpo de los números complejos. La matriz

$$P = \begin{bmatrix} -1 & 4 & 5 \\ 0 & 2 & -3 \\ 0 & 0 & 8 \end{bmatrix}$$

es inversible con inversa

$$P^{-1} = \begin{bmatrix} -1 & 2 & \frac{11}{8} \\ 0 & \frac{1}{2} & \frac{3}{16} \\ 0 & 0 & \frac{1}{8} \end{bmatrix}.$$

Así los vectores

$$\begin{aligned} \alpha_1' &= (-1, \ 0, 0) \\ \alpha_2' &= (\ 4, \ 2, 0) \\ \alpha_3' &= (\ 5, -3, 8) \end{aligned}$$

forman una base $\mathcal{B}'$ de F^3. Las coordenadas x_1', x_2', x_3' del vector $\alpha = (x_1, x_2, x_3)$ en la base $\mathcal{B}'$ vienen dadas por

$$\begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \begin{bmatrix} -x_1 + 2x_2 + \frac{11}{8}x_3 \\ \frac{1}{2}x_2 + \frac{3}{16}x_3 \\ \frac{1}{8}x_3 \end{bmatrix} = \begin{bmatrix} -1 & 2 & \frac{11}{8} \\ 0 & \frac{1}{2} & \frac{3}{16} \\ 0 & 0 & \frac{1}{8} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

En particular

$$(3, 2, -8) = -10\alpha_1' - \tfrac{1}{2}\alpha_2' - \alpha_3'.$$

Ejercicios

1. Demostrar que los vectores

$$\begin{aligned} \alpha_1 &= (1, 1, 0, 0), & \alpha_2 &= (0, 0, 1, 1) \\ \alpha_3 &= (1, 0, 0, 4), & \alpha_4 &= (0, 0, 0, 2) \end{aligned}$$

forman una base de R^4. Hallar las coordenadas de cada uno de los vectores de la base canónica respecto de la base ordenada $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$.

2. Hallar la matriz de coordenadas del vector (1, 0, 1) en la base de C^3 formada por los vectores $(2i, 1, 0)$, $(2, -1, 1)$, $(0, 1 + i, 1 - i)$, en ese orden.

3. Sea $\mathcal{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ la base ordenada de R^3 formada por

$$\alpha_1 = (1, 0, -1), \qquad \alpha_2 = (1, 1, 1), \qquad \alpha_3 = (1, 0, 0).$$

¿Cuáles son las coordenadas del vector (a, b, c) en la base ordenada $\mathcal{B}$?

4. Sea W el subespacio de C^3 generado por $\alpha_1 = (1, 0, i)$ y $\alpha_2 = (1 + i, 1, -1)$.

(a) Demostrar que α_1 y α_2 forman una base de W.

(b) Demostrar que los vectores $\beta_1 = (1, 1, 0)$ y $\beta_2 = (1, i, 1 + i)$ pertenecen a W y forman otra base de W.

(c) ¿Cuáles son las coordenadas de α_1 y α_2 en la base ordenada $\{\beta_1, \beta_2\}$ de W?

5. Sean $\alpha = (x_1, x_2)$ y $\beta = (y_1, y_2)$ dos vectores de R^2 tales que

$$x_1y_1 + x_2y_2 = 0, \qquad x_1^2 + x_2^2 = y_1^2 + y_2^2 = 1.$$

Demostrar que $\mathcal{B} = \{\alpha, \beta\}$ es una base de R^2. Hallar las coordenadas del vector (a, b) en la base ordenada $\mathcal{B} = \{\alpha, \beta\}$. (Las condiciones impuestas a α y β dicen, geométricamente, que α y β son perpendiculares y de longitud 1.)

6. Sea V el espacio vectorial sobre los números complejos de todas las funciones de R en C; es decir, el espacio de todas las funciones sobre el eje real a valor complejo. Sea $f_1(x) = 1$, $f_2(x) = e^{ix}$, $f_3(x) = e^{-ix}$.

(a) Demostrar que f_1, f_2 y f_3 son linealmente independientes.

(b) Sea $g_1(x) = 1$, $g_2(x) = \cos x$, $g_3(x) = \operatorname{sen} x$. Hallar una matriz inversible 3×3, P, tal que

$$g_j = \sum_{i=1}^{3} P_{ij} f_i.$$

7. Sea V el espacio vectorial (real) de todas las funciones polinomios de R en R de grado 2 o menor; es decir, el espacio de todas las funciones de la forma

$$f(x) = c_0 + c_1x + c_2x^2.$$

Sea t un número real fijo y defínase

$$g_1(x) = 1, \qquad g_2(x) = x + t, \qquad g_3(x) = (x + t)^2.$$

Demostrar que $\mathcal{B} = \{g_1, g_2, g_3\}$ es una base de V. Si

$$f(x) = c_0 + c_1x + c_2x^2$$

¿Cuáles son las coordenadas de f en esta base ordenada $\mathcal{B}$?

2.5. *Resumen de equivalencia por filas*

En esta sección se usarán unos hechos elementales referentes a bases y dimensión en espacios vectoriales de dimensión finita para completar el tratamiento de la equivalencia de matrices por filas. Se recuerda que si A es una matriz $m \times n$ sobre el cuerpo F, los vectores fila de A son los vectores $\alpha_1, \ldots, \alpha_m$ de F^n definidos por

$$\alpha_i = (A_{i1}, \ldots, A_{in}).$$

y que el espacio de filas de A es el subespacio de F^n generado por estos vectores. El **rango de fila** de A es la dimensión del espacio de filas de A.

Si P es una matriz $k \times m$ sobre F, entonces el producto $B = PA$ es una matriz $k \times n$ cuyos vectores fila $\beta_1, \ldots, \beta_k$ son combinaciones lineales

$$\beta_i = P_{i1}\alpha_1 + \cdots + P_{im}\alpha_m$$

de los vectores fila de A. Así el espacio de filas de B es un subespacio del espacio de filas de A. Si P es una matriz inversible $m \times m$, entonces B es equivalente por filas a A, de modo que la simetría de la equivalencia por filas, o la igualdad $A = P^{-1}B$, indica que el espacio de filas de A es también un subespacio del espacio de filas de B.

Teorema 9. *Las matrices equivalentes por filas tienen el mismo espacio de filas.*

De este modo se ha visto que para estudiar el espacio de filas de A se puede estudiar el espacio de filas de una matriz escalón reducida por filas que es equivalente por filas a A, como se hará en seguida.

Teorema 10. *Sea R una matriz escalón no nula reducida por filas. Entonces los vectores fila no nulos de R forman una base del espacio de filas de R.*

Demostración. Sean $\rho_1, \ldots, \rho_r$ los vectores fila no nulos de R

$$\rho_i = (R_{i1}, \ldots, R_{in}).$$

Ciertamente estos vectores generan el espacio de filas de R; se necesita solo demostrar que son linealmente independientes. Como R es una matriz escalón reducida por filas, hay enteros positivos $k_1, \ldots, k_r$ tales que, para $i < r$

(2-18)
$$\begin{aligned} &\text{(a) } R(i, j) = 0 \quad \text{si} \quad j < k_i \\ &\text{(b) } R(i, k_j) = \delta_{ij} \\ &\text{(c) } k_1 < \cdots < k_r. \end{aligned}$$

Supóngase que $\beta = (b_1, \ldots, b_n)$ sea un vector del espacio de filas de R:

(2-19)
$$\beta = c_1\rho_1 + \cdots + c_r\rho_r.$$

Se afirma entonces que $c_j = b_{k_j}$. En efecto, por (2-18)

(2-20)
$$\begin{aligned} b_{k_j} &= \sum_{i=1}^{r} c_i R(i, k_j) \\ &= \sum_{i=1}^{r} c_i \delta_{ij} \\ &= c_j. \end{aligned}$$

En particular, si $\beta = 0$, es decir, si $c_1\rho_1 + \cdots + c_r\rho_r = 0$, entonces c_j debe ser la k-ésima coordenada del vector nulo, con lo que $c_j = 0$, $j = 1, \ldots, r$. Así, pues, $\rho_1, \ldots, \rho_r$ son linealmente independientes. ∎

Teorema 11. *Sean m y n dos enteros positivos y sea F un cuerpo. Supóngase que W es un subespacio de F^n y que dim $W \leq m$. Entonces existe exactamente una sola matriz escalón $m \times n$ reducida por filas sobre F que tiene a W como su espacio de filas.*

Demostración. Hay al menos una matriz escalón reducida por fila $m \times n$ con espacio de filas W. Como dim $W \leq m$, se pueden elegir m vectores $\alpha_1, \ldots, \alpha_m$ de W que generan W. Sea A la matriz $m \times n$ con vectores fila $\alpha_1, \ldots, \alpha_m$ y sea R la matriz escalón reducida por filas que es equivalente por filas a A. Entonces, el espacio de filas de R es W.

Sea ahora R cualquier matriz escalón reducida por filas que tiene a W como su espacio de filas. Sean $\rho_1, \ldots, \rho_r$ los vectores fila no nulos de R y supóngase que el elemento principal no nulo de ρ_i esté en la columna k_i, $i = 1, \ldots, r$. Los vectores $\rho_1, \ldots, \rho_r$ forman una base de W. En la demostración del Teorema 10 se observó que si $\beta = (b_1, \ldots, b_n)$ pertenece a W, entonces

$$\beta = c_1\rho_1 + \cdots + c_r\rho_r,$$

y $c_i = b_{k_i}$; en otras palabras, la expresión única de β como combinación lineal de los $\rho_1, \ldots, \rho_r$ es

$$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i. \tag{2-21}$$

Así, pues, cada vector β está determinado si se conocen las coordenadas b_{k_i}, $i = 1, \ldots, r$. Por ejemplo, ρ_s es el vector único de W que tiene la k_s-ésima coordenada 1 y las k_i-ésimas coordenadas 0 para $i \neq s$.

Supóngase que β pertenece a W y que $\beta \neq 0$. Se afirma que la primera coordenada no nula de β está en una de las columnas k_s. Como

$$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i$$

y $\beta \neq 0$ se puede escribir

$$\beta = \sum_{i=s}^{r} b_{k_i}\rho_i, \qquad b_{k_s} \neq 0. \tag{2-22}$$

De las condiciones (2-18) se tiene $R_{ij} = 0$ si $i > s$ y $j \geq k_s$. Con lo que

$$\beta = (0, \ldots, 0, \;\; b_{k_s}, \ldots, b_n), \qquad b_{k_s} \neq 0$$

y la primera coordenada no nula de β está en la columna k_s. Obsérvese también que para cada k_s, $s = 1, \ldots, r$, existe un vector en W que tiene su k_s-ésima coordenada no nula, a saber ρ_s.

Ahora es claro que R está determinada unívocamente por W. La descripción de R en términos de W es como sigue. Se consideran todos los vectores $\beta = (b_1, \ldots, b_n)$ en W. Si $\beta \neq 0$, entonces la primera coordenada no nula debe estar en alguna columna t:

$$\beta = (0, \ldots, 0, \;\; b_t, \ldots, b_n), \qquad b_t \neq 0.$$

Sean $k_1, \ldots, k_r$ aquellos enteros positivos t tales que existe algún $\beta \neq 0$ en W,

cuya primera coordenada no nula está en la columna t. Se disponen los $k_1, \ldots, k_r$ en el orden $k_1 < k_2 < \cdots < k_r$. Para cada uno de los enteros positivos k_s habrá un vector ρ_s y solo uno de W tal que la k_s-ésima coordenada de ρ_s es 1 y la k_i-ésima es 0 para $i \neq s$. Entonces R ès la matriz $m \times n$ que tiene los vectores fila $\rho_1, \ldots, \rho_r, 0, \ldots, 0$. ▌

Corolario. *Cada matriz $m \times n$, A, es equivalente por filas a una, y solamente una, matriz escalón reducida por filas.*

Demostración. Sabemos que A es equivalente por filas, al menos, a una matriz escalón reducida por filas. Si A es equivalente por filas a otra matriz escalón R', entonces R es equivalente por filas a R'; luego, R y R' tienen el mismo espacio de filas y deben ser idénticas. ▌

Corolario. *Sean A y B matrices $m \times n$ sobre el cuerpo F. Entonces A y B son equivalentes por filas si, y solo si, tienen el mismo espacio de filas.*

Demostración. Se sabe que si A y B son equivalentes por filas, entonces tienen el mismo espacio de filas. Supóngase ahora que A y B tengan el mismo espacio de filas. Ahora A es equivalente por filas a una matriz escalón reducida por filas R, y B es equivalente por filas a una matriz reducida por filas R'. Como A y B tienen el mismo espacio de filas, R y R' tienen el mismo espacio de filas. Con lo que $R = R'$ y A es equivalente por filas a B. ▌

En resumen, si A y B son dos matrices $m \times n$ sobre el cuerpo F, las siguientes afirmaciones son equivalentes:

1. A y B son equivalentes por filas.
2. A y B tienen el mismo espacio de filas.
3. $B \doteq PA$, donde P es una matriz inversible $m \times m$.

Una cuarta afirmación equivalente es que los sistemas homogéneos $AX = 0$ y $BX = 0$ tienen las mismas soluciones; sin embargo, aun cuando se sabe que la equivalencia por filas de A y B implica que estos sistemas tienen las mismas soluciones, es más conveniente dejar la demostración del recíproco para más adelante.

2.6. *Cálculos relativos a subespacios*

Queremos ahora mostrar cómo las operaciones elementales por filas dan un método normalizado para responder a ciertas preguntas concretas concernientes a subespacios de F^n. Ya hemos deducido lo que vamos a necesitar. El análisis es válido para cualquier espacio vectorial V de dimensión n sobre el cuerpo F, si se elige una base ordenada $\mathcal{B}$ fija y se expresa cada vector α de V por la n-tuple $(x_1, \ldots, x_n)$ que da las coordenadas de α en la base ordenada $\mathcal{B}$.

Supóngase que se han dado m vectores $\alpha_1, \ldots, \alpha_m$ pertenecientes a F^n. Consideremos las siguientes preguntas:

1. ¿Cómo se determina si los vectores $\alpha_1, \ldots, \alpha_m$ son linealmente independientes? O en términos generales, ¿cómo se determina la dimensión del subespacio W generado por estos vectores?

2. Dado β en F^n, ¿cómo se determina si β es una combinación lineal de $\alpha_1, \ldots, \alpha_m$, es decir, si β pertenece al subespacio W?
3. ¿Cómo se puede dar una descripción explícita del subespacio W?

La tercera pregunta es algo vaga, ya que no se especifica qué se entiende por una «descripción explícita»; sin embargo, ello se aclarará cuando se dé el tipo de descripción que se tiene en mente. Con esta descripción las preguntas (1) y (2) pueden contestarse inmediatamente.

Sea A una matriz $m \times n$ con vectores fila α_i:

$$\alpha_i = (A_{i1}, \ldots, A_{in})$$

Hágase una sucesión de operaciones elementales por fila, empezando con A para terminar con una matriz escalón reducida por fila R. Anteriormente ya se indicó cómo se hace esto; ahora, la dimensión de W (el espacio de filas de A) queda de manifiesto, ya que esta dimensión es simplemente el número de vectores fila no nulos de R. Si $\rho_1, \ldots, \rho_r$ son los vectores fila no nulos de R, entonces $\mathfrak{B} = \{\rho_1, \ldots, \rho_r\}$ es una base de W. Si la primera coordenada no nula de ρ_i es la coordenada k_i-ésima, se tiene entonces para $i \leq r$

(a) $$R(i, j) = 0, \quad \text{si} \quad j < k_i$$
(b) $$R(i, k_j) = \delta_{ij}$$
(c) $$k_1 < \cdots < k_r.$$

El subespacio W consta de todos los vectores

$$\beta = c_1\rho_1 + \cdots + c_r\rho_r$$
$$= \sum_{i=1}^{r} c_i(R_{i1}, \ldots, R_{in}).$$

Las coordenadas $b_1, \ldots, b_n$ de tal vector β, son entonces

(2-23) $$b_j = \sum_{i=1}^{r} c_i R_{ij}.$$

En particular, $b_{k_j} = c_j$, y así, si $\beta = (b_1, \ldots, b_n)$ es una combinación lineal de los ρ_i, debe ser la combinación lineal particular

(2-24) $$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i.$$

Las condiciones sobre β que (2-24) ha de tener son

(2-25) $$b_j = \sum_{i=1}^{r} b_{k_i} R_{ij}, \qquad j = 1, \ldots, n.$$

Ahora (2-25) es la descripción explícita del subespacio W generado por $\alpha_1, \ldots, \alpha_m$, esto es, el subespacio consta de todos los vectores β de F^n cuyas coordenadas satisfacen (2-25). ¿Qué tipo de descripción es (2-25)? En primer lugar describe W como todas las soluciones $\beta = (b_1, \ldots, b_n)$ del sistema de ecuaciones lineales homogéneas (2-25). Este sistema de ecuaciones es de una naturaleza muy especial, porque expresa $(n - r)$ de las coordenadas como combinaciones lineales de las r coordenadas señaladas, $b_{k_1}, \ldots, b_{k_r}$. Se tiene

completa libertad para elegir las coordenadas b_k; esto es, si $c_1, \ldots, c_r$ son r escalares cualesquiera, existe un vector β de W y solo uno que tiene c_i como su k_i-ésima coordenada.

Lo significativo, aquí, es que: dados los vectores α_i, la reducción por fila es un método directo para determinar los enteros $r, k_1, \ldots, k_r$ y los escalares R_{ij} que dan la descripción (2-25) del subespacio generado por $\alpha_1, \ldots, \alpha_m$. Debemos observar, como hicimos en el Teorema 11, que todo subespacio W de F^n tiene una descripción del tipo (2-25). También debemos hacer notar algunos aspectos respecto a la pregunta (2). Se ha determinado, en la Sección 1.4, cómo se puede hallar una matriz inversible $m \times m$, P, tal que $R = PA$. El conocimiento de P permite encontrar los escalares $x_1, \ldots, x_m$ tales que

$$\beta = x_1\alpha_1 + \cdots + x_m\alpha_m$$

cuando ello es posible. En efecto, los vectores fila de R están dados por

$$\rho_i = \sum_{j=1}^{m} P_{ij}\alpha_j$$

de modo que si β es una combinación lineal de los α_i, se tiene

$$\begin{aligned}\beta &= \sum_{i=1}^{r} b_{k_i}\rho_i \\ &= \sum_{i=1}^{r} b_{k_i} \sum_{j=1}^{m} P_{ij}\alpha_j \\ &= \sum_{j=1}^{m} \sum_{i=1}^{r} b_{k_i}P_{ij}\alpha_j\end{aligned}$$

con lo que

$$x_j = \sum_{i=1}^{r} b_{k_i}P_{ij}$$

es una posible elección de los x_j (puede haber varias).

La cuestión de si $\beta = (b_1, \ldots, b_n)$ es una combinación lineal de las α_i, y en tal caso cuáles son los escalares x_i se puede también considerar preguntándose si el sistema de ecuaciones

$$\sum_{i=1}^{m} A_{ij}x_i = b_j, \qquad j = 1, \ldots, n$$

tiene solución y cuáles son las soluciones. La matriz coeficiente de este sistema es la matriz $n \times m$, B, con vectores *columna* $\alpha_1, \ldots, \alpha_m$. En el Capítulo 1 se estudió el uso de las operaciones elementales por fila para resolver un sistema de ecuaciones $BX = Y$. Consideremos un ejemplo en el cual adoptamos ambos puntos de vista para responder preguntas respecto a subespacios de F^n.

Ejemplo 21. Consideremos el siguiente problema. Sea W el subespacio de R^4 generado por los vectores

$$\begin{aligned}\alpha_1 &= (1, 2, 2, 1)\\ \alpha_2 &= (0, 2, 0, 1)\\ \alpha_3 &= (-2, 0, -4, 3).\end{aligned}$$

(a) Demostrar que α_1, α_2, α_3 forman una base de W, es decir, que estos vectores son linealmente independientes.

(b) Sea $\beta = (b_1, b_2, b_3, b_4)$ un vector de W. ¿Cuáles son las coordenadas de β respecto a la base ordenada $\{\alpha_1, \alpha_2, \alpha_3\}$?

(c) Sean

$$\begin{aligned} \alpha_1' &= (1, 0, 2, 0) \\ \alpha_2' &= (0, 2, 0, 1) \\ \alpha_3' &= (0, 0, 0, 3). \end{aligned}$$

Demostrar que α_1', α_2', α_3' forman una base de W.

(d) Si β pertenece a W, se designa por X la matriz de coordenadas de β respecto a la base de los α, y por X' a la matriz de coordenadas de β respecto a la base de los α'. Hallar la matriz 3×3, P, tal que $X = PX'$ para cada una de tales β.

Para responder estas preguntas por el primer método se forma la matriz A con vectores fila α_1, α_2, α_3, se determina la matriz escalón reducida por filas R, que es equivalente por filas a A, y haciendo las mismas operaciones sobre la matriz identidad se obtiene la matriz inversible Q, tal que $R = QA$:

$$\begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ -2 & 0 & -4 & 3 \end{bmatrix} \rightarrow R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow Q = \tfrac{1}{6} \begin{bmatrix} 6 & -6 & 0 \\ -2 & 5 & -1 \\ 4 & -4 & 2 \end{bmatrix}$$

(a) Evidentemente R es de rango 3, con lo que α_1, α_2, α_3 son independientes.

(b) ¿Qué vectores $\beta = (b_1, b_2, b_3, b_4)$ pertenecen a W? Se tiene la base de W dada por ρ_1, ρ_2, ρ_3, que son los vectores fila de R. A la vista está que el espacio generado por ρ_1, ρ_2, ρ_3 consta de los vectores β para los que $b_3 = 2b_1$. Para un β tal, se tiene

$$\begin{aligned} \beta &= b_1\rho_1 + b_2\rho_2 + b_4\rho_3 \\ &= [b_1, b_2, b_4]R \\ &= [b_1 \quad b_2 \quad b_4]QA \\ &= x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3 \end{aligned}$$

donde $x_i = [b_1 \quad b_2 \quad b_3]Q_i$:

$$\begin{aligned} x_1 &= b_1 - \tfrac{1}{3}b_2 + \tfrac{2}{3}b_4 \\ x_2 &= -b_1 + \tfrac{5}{6}b_2 - \tfrac{2}{3}b_4 \\ x_3 &= - \tfrac{1}{6}b_2 + \tfrac{1}{3}b_4. \end{aligned} \tag{2-26}$$

(c) Los vectores α_1', α_2', α_3' son todos de la forma (y_1, y_2, y_3, y_4) con $y_3 = 2y_1$, y, por tanto, pertenecen a W. Es evidente que son linealmente independientes.

(d) La matriz P tiene por columnas

$$P_j = [\alpha_j']_{\mathfrak{B}}$$

donde $\mathfrak{B} = \{\alpha_1, \alpha_2, \alpha_3\}$. Las ecuaciones (2-26 dicen cómo hallar la matriz de coordenadas para $\alpha_1', \alpha_2', \alpha_3'$. Por ejemplo, con $\beta = \alpha_1'$ se tiene $b_1 = 1$, $b_2 = 0$, $b_3 = 2$, $b_4 = 0$, y

$$\begin{aligned} x_1 &= \quad 1 - \tfrac{1}{3}(0) + \tfrac{2}{3}(0) = \quad 1 \\ x_2 &= -1 + \tfrac{5}{6}(0) - \tfrac{2}{3}(0) = -1 \\ x_3 &= \quad\;\; - \tfrac{1}{6}(0) + \tfrac{1}{3}(0) = \quad 0. \end{aligned}$$

Con lo que $\alpha_1' = \alpha_1 - \alpha_2$. En forma semejante, se obtiene $\alpha_2' = \alpha_2$ y $\alpha_3' = 2\alpha_1 - 2\alpha_2 + \alpha_3$. Luego

$$P = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}.$$

Ahora se verá cómo se responden las preguntas por el segundo método descrito. Se forma la matriz 4×3, B, con vectores columna $\alpha_1, \alpha_2, \alpha_3$:

$$B = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 2 & 0 \\ 2 & 0 & -4 \\ 1 & 1 & 3 \end{bmatrix}.$$

Inquirimos para cuáles y_1, y_2, y_3, y_4 el sistema $BX = Y$ tiene una solución.

$$\begin{bmatrix} 1 & 0 & -2 & y_1 \\ 2 & 2 & 0 & y_2 \\ 2 & 0 & -4 & y_3 \\ 1 & 1 & 3 & y_4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -2 & y_1 \\ 0 & 2 & 4 & y_2 - 2y_1 \\ 0 & 0 & 0 & y_3 - 2y_1 \\ 0 & 1 & 5 & y_4 - y_1 \end{bmatrix} \longrightarrow$$

$$\begin{bmatrix} 1 & 0 & -2 & y_1 \\ 0 & 0 & -6 & y_2 - 2y_4 \\ 0 & 1 & 5 & y_4 - y_1 \\ 0 & 0 & 0 & y_3 - 2y_1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & y_1 - \tfrac{1}{3}y_2 + \tfrac{2}{3}y_4 \\ 0 & 0 & 1 & \tfrac{1}{6}(2y_4 - y_2) \\ 0 & 1 & 0 & -y_1 + \tfrac{5}{6}y_2 - \tfrac{2}{3}y_4 \\ 0 & 0 & 0 & y_3 - 2y_1 \end{bmatrix}$$

Así la condición para que el sistema $BX = Y$ tenga una solución es que $y_3 = 2y_1$. Así $\beta = (b_1, b_2, b_3, b_4)$ pertenece a W si y, solo si, $b_3 = 2b_1$. Si β pertenece a W, entonces las coordenadas (x_1, x_2, x_3) en la base ordenada $\{\alpha_1, \alpha_2, \alpha_3\}$ se pueden leer de la última matriz anterior. Obtenemos nuevamente las fórmulas (2-26) para esas coordenadas.

Las preguntas (c) y (d) se responden como antes.

Ejemplo 22. Consideremos la matriz 5×5

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

y los siguientes problemas concernientes a A.

(a) Hallar una matriz inversible P, tal que PA sea una matriz escalón reducida por filas R.

(b) Hallar una base para el espacio de filas W de A.

(c) Determinar qué vectores $(b_1, b_2, b_3, b_4, b_5)$ están en W.

(d) Hallar la matriz de coordenadas de cada vector $(b_1, b_2, b_3, b_4, b_5)$ de W en la base ordenada elegida en (b).

(e) Escribir cada vector $(b_1, b_2, b_3, b_4, b_5)$ de W como una combinación lineal de las filas de A.

(f) Dar una descripción explícita del espacio vectorial V de todas las matrices columnas de 5×1, X, tales que $AX = 0$.

(g) Hallar una base de V.

(h) ¿Para qué matriz columna 5×1, Y, la ecuación $AX = Y$ tiene soluciones X?

Para resolver estos problemas se forma la matriz aumentada A' del sistema $AX = Y$ y efectuando una sucesión apropiada de operaciones por filas sobre A':

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 1 & 2 & -1 & -1 & 0 & y_2 \\ 0 & 0 & 1 & 4 & 0 & y_3 \\ 2 & 4 & 1 & 10 & 1 & y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & -1 & -4 & 0 & -y_1 + y_2 \\ 0 & 0 & 1 & 4 & 0 & y_3 \\ 0 & 0 & 1 & 4 & 1 & -2y_1 + y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \longrightarrow$$

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & 1 & 4 & 0 & y_1 - y_2 \\ 0 & 0 & 0 & 0 & 0 & -y_1 + y_2 + y_3 \\ 0 & 0 & 0 & 0 & 1 & -3y_1 + y_2 + y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \longrightarrow$$

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & 1 & 4 & 0 & y_1 - y_2 \\ 0 & 0 & 0 & 0 & 1 & y_5 \\ 0 & 0 & 0 & 0 & 0 & -y_1 + y_2 + y_3 \\ 0 & 0 & 0 & 0 & 0 & -3y_1 + y_2 + y_4 - y_5 \end{bmatrix}$$

(a) Si

$$PY = \begin{bmatrix} y_1 \\ y_1 - y_2 \\ y_5 \\ -y_1 + y_2 + y_3 \\ -3y_1 + y_2 + y_4 - y_5 \end{bmatrix}$$

para todo Y, entonces

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 \\ -3 & 1 & 0 & 1 & -1 \end{bmatrix}$$

luego PA es la matriz escalón reducida por filas

$$R = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Se debe resaltar que la matriz P no es única. Existen, de hecho, muchas matrices inversibles P (que provienen de los diferentes modos de elegir las operaciones usadas para reducir A') tales que $PA = R$.

(b) Una base de W se puede tener al tomar las filas no nulas

$$\begin{aligned} \rho_1 &= (1 \quad 2 \quad 0 \quad 3 \quad 0) \\ \rho_2 &= (0 \quad 0 \quad 1 \quad 4 \quad 0) \\ \rho_3 &= (0 \quad 0 \quad 0 \quad 0 \quad 1) \end{aligned}$$

de R.

(c) El espacio de filas W consta de todos los vectores de la forma

$$\begin{aligned} \beta &= c_1\rho_1 + c_2\rho_2 + c_3\rho_3 \\ &= (c_1, 2c_1, c_2, 3c_1 + 4c_2, c_3) \end{aligned}$$

donde c_1, c_2, c_3 son escalares arbitrarios. Así, $(b_1, b_2, b_3, b_4, b_5)$ pertenece a W si y, solo si,

$$(b_1, b_2, b_3, b_4, b_5) = b_1\rho_1 + b_3\rho_2 + b_5\rho_3$$

lo que es cierto si y, solo si,

$$\begin{aligned} b_2 &= 2b_1 \\ b_4 &= 3b_1 + 4b_3. \end{aligned}$$

Estas ecuaciones son casos particulares del sistema general (2-25), que se puede usar para ver inmediatamente si un vector dado pertenece a W. Así, $(-5, -10, 1, -11, 20)$ es una combinación lineal de las filas de A, pero $(1, 2, 3, 4, 5)$ no lo es.

(d) La matriz de coordenadas del vector $(b_1, 2b_1, b_3, 3b_1 + 4b_3, b_5)$ en la base ρ_1 ρ_2, ρ_3 es evidentemente

$$\begin{bmatrix} b_1 \\ b_3 \\ b_5 \end{bmatrix}.$$

(e) Hay varias maneras de escribir los vectores de W como combinaciones lineales de las filas de A. Tal vez el método más fácil es seguir el primer procedimiento del Ejemplo 21 anterior:

$$\begin{aligned} \beta &= (b_1, 2b_1, b_3, 3b_1 + 4b_3, b_5) \\ &= [b_1, b_3, b_5, 0, 0] \cdot R \\ &= [b_1, b_3, b_5, 0, 0] \cdot PA \\ &= [b_1, b_3, b_5, 0, 0] \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 \\ -3 & 1 & 0 & 1 & -1 \end{bmatrix} \cdot A \\ &= [b_1 + b_3, -b_3, 0, 0, b_5] \cdot A. \end{aligned}$$

En particular, con $\beta = (-5, -10, 1, -11, 20)$ se tiene

$$\beta = (-4, -1, 0, 0, 20)\begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

(f) Las ecuaciones en el sistema $RX = 0$ son:

$$\begin{aligned} x_1 + 2x_2 + 3x_4 &= 0 \\ x_3 + 4x_4 &= 0 \\ x_5 &= 0. \end{aligned}$$

Con lo que V consta de todas las columnas de la forma

$$X = \begin{bmatrix} -2x_2 - 3x_4 \\ x_2 \\ -4x_4 \\ x_4 \\ 0 \end{bmatrix}$$

donde x_2 y x_4 son arbitrarios.

(g) Las columnas

$$\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \\ 0 \end{bmatrix}$$

forman una base para V. Este es un ejemplo de la base descrita en el Ejemplo 15.

(h) La ecuación $AX = Y$ tiene las soluciones X si y, solo si,

$$\begin{aligned} -y_1 + y_2 + y_3 \qquad &= 0 \\ -3y_1 + y_2 + y_4 - y_5 &= 0. \end{aligned}$$

Ejercicios

1. Sea $s < n$ y A una matriz $s \times n$ con elementos en el cuerpo F. Usando el Teorema 4 (no su demostración), demostrar que existe un X no nulo en $F^{n \times 1}$ tal que $AX = 0$.

2. Sean

$$\alpha_1 = (1, 1, -2, 1), \qquad \alpha_2 = (3, 0, 4, -1), \qquad \alpha_3 = (-1, 2, 5, 2).$$

y sean

$$\alpha = (4, -5, 9, -7), \qquad \beta = (3, 1, -4, 4), \qquad \gamma = (-1, 1, 0, 1).$$

(a) ¿Cuál de los vectores α, β, γ pertenece al subespacio R^4 generado por los α_i?
(b) ¿Cuáles de los vectores α, β, γ están en el subespacio de C^4 generado por los α_i?
(c) ¿Sugiere esto algún teorema?

luego PA es la matriz escalón reducida por filas

$$R = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Se debe resaltar que la matriz P no es única. Existen, de hecho, muchas matrices inversibles P (que provienen de los diferentes modos de elegir las operaciones usadas para reducir A') tales que $PA = R$.

(b) Una base de W se puede tener al tomar las filas no nulas

$$\begin{aligned} \rho_1 &= (1 \quad 2 \quad 0 \quad 3 \quad 0) \\ \rho_2 &= (0 \quad 0 \quad 1 \quad 4 \quad 0) \\ \rho_3 &= (0 \quad 0 \quad 0 \quad 0 \quad 1) \end{aligned}$$

de R.

(c) El espacio de filas W consta de todos los vectores de la forma

$$\begin{aligned} \beta &= c_1\rho_1 + c_2\rho_2 + c_3\rho_3 \\ &= (c_1, 2c_1, c_2, 3c_1 + 4c_2, c_3) \end{aligned}$$

donde c_1, c_2, c_3 son escalares arbitrarios. Así, $(b_1, b_2, b_3, b_4, b_5)$ pertenece a W si y, solo si,

$$(b_1, b_2, b_3, b_4, b_5) = b_1\rho_1 + b_3\rho_2 + b_5\rho_3$$

lo que es cierto si y, solo si,

$$\begin{aligned} b_2 &= 2b_1 \\ b_4 &= 3b_1 + 4b_3. \end{aligned}$$

Estas ecuaciones son casos particulares del sistema general (2-25), que se puede usar para ver inmediatamente si un vector dado pertenece a W. Así, $(-5, -10, 1, -11, 20)$ es una combinación lineal de las filas de A, pero $(1, 2, 3, 4, 5)$ no lo es.

(d) La matriz de coordenadas del vector $(b_1, 2b_1, b_3, 3b_1 + 4b_3, b_5)$ en la base ρ_1 ρ_2, ρ_3 es evidentemente

$$\begin{bmatrix} b_1 \\ b_3 \\ b_5 \end{bmatrix}.$$

(e) Hay varias maneras de escribir los vectores de W como combinaciones lineales de las filas de A. Tal vez el método más fácil es seguir el primer procedimiento del Ejemplo 21 anterior:

$$\begin{aligned} \beta &= (b_1, 2b_1, b_3, 3b_1 + 4b_3, b_5) \\ &= [b_1, b_3, b_5, 0, 0] \cdot R \\ &= [b_1, b_3, b_5, 0, 0] \cdot PA \\ &= [b_1, b_3, b_5, 0, 0] \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 \\ -3 & 1 & 0 & 1 & -1 \end{bmatrix} \cdot A \\ &= [b_1 + b_3, -b_3, 0, 0, b_5] \cdot A. \end{aligned}$$

En particular, con $\beta = (-5, -10, 1, -11, 20)$ se tiene

$$\beta = (-4, -1, 0, 0, 20)\begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

(f) Las ecuaciones en el sistema $RX = 0$ son:

$$\begin{aligned} x_1 + 2x_2 + 3x_4 &= 0 \\ x_3 + 4x_4 &= 0 \\ x_5 &= 0. \end{aligned}$$

Con lo que V consta de todas las columnas de la forma

$$X = \begin{bmatrix} -2x_2 - 3x_4 \\ x_2 \\ -4x_4 \\ x_4 \\ 0 \end{bmatrix}$$

donde x_2 y x_4 son arbitrarios.

(g) Las columnas

$$\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \\ 0 \end{bmatrix}$$

forman una base para V. Este es un ejemplo de la base descrita en el Ejemplo 15.

(h) La ecuación $AX = Y$ tiene las soluciones X si y, solo si,

$$\begin{aligned} -y_1 + y_2 + y_3 \qquad &= 0 \\ -3y_1 + y_2 + y_4 - y_5 &= 0. \end{aligned}$$

Ejercicios

1. Sea $s < n$ y A una matriz $s \times n$ con elementos en el cuerpo F. Usando el Teorema 4 (no su demostración), demostrar que existe un X no nulo en $F^{n \times 1}$ tal que $AX = 0$.

2. Sean

$$\alpha_1 = (1, 1, -2, 1), \qquad \alpha_2 = (3, 0, 4, -1), \qquad \alpha_3 = (-1, 2, 5, 2).$$

y sean

$$\alpha = (4, -5, 9, -7), \qquad \beta = (3, 1, -4, 4), \qquad \gamma = (-1, 1, 0, 1).$$

(a) ¿Cuál de los vectores α, β, γ pertenece al subespacio R^4 generado por los α_i?
(b) ¿Cuáles de los vectores α, β, γ están en el subespacio de C^4 generado por los α_i?
(c) ¿Sugiere esto algún teorema?

3. Considerar los vectores en R^4 definidos por

$$\alpha_1 = (-1, 0, 1, 2), \qquad \alpha_2 = (3, 4, -2, 5), \qquad \alpha_3 = (1, 4, 0, 9).$$

Hallar un sistema de ecuaciones lineales homogéneas para las que el espacio de las soluciones sea exactamente el subespacio de R^4 generado por los tres vectores dados.

4. En C^3, sean

$$\alpha_1 = (1, 0, -i), \qquad \alpha_2 = (1 + i, 1 - i, 1), \qquad \alpha_3 = (i, i, i).$$

Demostrar que estos vectores forman una base de C^3. ¿Cuáles son las coordenadas del vector (a, b, c) en esta base?

5. Dar una descripción explícita del tipo (2-25) para los vectores

$$\beta = (b_1, b_2, b_3, b_4, b_5)$$

de R^5 que son combinaciones lineales de los vectores

$$\begin{aligned} \alpha_1 &= (1, 0, 2, 1, -1), \qquad & \alpha_2 &= (-1, 2, -4, 2, 0) \\ \alpha_3 &= (2, -1, 5, 2, 1), \qquad & \alpha_4 &= (2, 1, 3, 5, 2). \end{aligned}$$

6. Sea V un espacio vectorial real generado por las filas de la matriz

$$A = \begin{bmatrix} 3 & 21 & 0 & 9 & 0 \\ 1 & 7 & -1 & -2 & -1 \\ 2 & 14 & 0 & 6 & 1 \\ 6 & 42 & -1 & 13 & 0 \end{bmatrix}.$$

(a) Hallar una base para V.

(b) ¿Qué vectores $(x_1, x_2, x_3, x_4, x_5)$ son elementos de V?

(c) Si $(x_1, x_2, x_3, x_4, x_5)$ pertenece a V, ¿cuáles son sus coordenadas en la base elegida en la parte (a)?

7. Sea A una matriz $m \times n$ sobre el cuerpo F y considerar el sistema de ecuaciones $AX = Y$. Demostrar que este sistema de ecuaciones tiene una solución si, y solo si, el rango de fila de A es igual al rango de fila de la matriz aumentada del sistema.

3. Transformaciones lineales

3.1. Transformaciones lineales

Vamos a introducir ahora las transformaciones lineales, que es de lo que se trata en la mayor parte de lo que resta del libro. El lector encontrará de utilidad leer (o releer) el estudio sobre funciones en el Apéndice, ya que se hará amplio uso de la terminología pertinente.

Definición. *Sean V y W dos espacios vectoriales sobre el cuerpo F. Una* **transformación lineal** *de V en W es una función T de V en W tal que*

$$T(c\alpha + \beta) = c(T\alpha) + T\beta$$

para todos los vectores α y β de V y todos los escalares c de F.

Ejemplo 1. Si V es cualquier espacio vectorial, la **transformación identidad** I, definida por $I\alpha = \alpha$, es una transformación lineal de V en V. La **transformación cero** 0, definida por $0\alpha = 0$, es una transformación lineal de V en V.

Ejemplo 2. Sea F un cuerpo y sea V el espacio vectorial de las funciones polinomios f de F en F, dado por

$$f(x) = c_0 + c_1x + \cdots + c_kx^k.$$

Sea

$$(Df)(x) = c_1 + 2c_2x + \cdots + kc_kx^{k-1}.$$

Entonces D es una transformación lineal de V en V: la transformación derivación.

Ejemplo 3. Sea A una matriz $m \times n$ dada, con elementos en el cuerpo F. La función T definida por $T(X) = AX$ es una transformación lineal de $F^{n\times 1}$ en $F^{m\times 1}$. La función U definida por $U(\alpha) = \alpha A$ es una transformación lineal de F^m en F^n.

Ejemplo 4. Sea P una matriz $m \times m$ dada, con elementos en el cuerpo F, y sea Q otra matriz $n \times n$ dada, sobre F. Se define una función T del espacio $F^{m\times n}$ en sí mismo por $T(A) = PAQ$. Entonces T es una transformación lineal de $F^{m\times n}$ en $F^{m\times n}$, porque

$$\begin{aligned} T(cA + B) &= P(cA + B)Q \\ &= (cPA + PB)Q \\ &= cPAQ + PBQ \\ &= cT(A) + T(B). \end{aligned}$$

Ejemplo 5. Sea R el cuerpo de los números reales y sea V el espacio de todas las funciones continuas de R en R. Se define T por

$$(Tf)(x) = \int_0^x f(t)\, dt.$$

Entonces T es una transformación lineal de V en V. La función Tf no solo es continua, sino que también tiene primera derivada continua. La linealidad de la integración es una de sus propiedades fundamentales.

El lector no tendrá dificultades en verificar que las transformaciones definidas en los Ejemplos 1, 2, 3 y 5 son transformaciones lineales. La lista de ejemplos se ampliará considerablemente cuando se estudien más aspectos de las transformaciones lineales.

Es importante observar que si T es una transformación lineal de V en W, entonces $T(0) = 0$; lo que se ve por la misma definición, pues

$$T(0) = T(0 + 0) = T(0) + T(0).$$

Este punto es a menudo motivo de confusión para quien estudia álgebra lineal por primera vez, ya que probablemente haya tenido una visión un poco diferente del uso del término «función lineal». Un breve comentario al respecto podría aclarar esa confusión. Supóngase que V es el espacio vectorial R^1. Una transformación lineal de V en V es entonces un tipo especial de función real en el eje real. En un curso de cálculo es probable que se diga que tal función es función lineal si su grafo es una recta. Una transformación lineal de R^1 en R^1, de acuerdo con la definición, será una función de R en R cuyo grafo es una recta que *pasa por el origen.*

A más de la propiedad $T(0) = 0$ indiquemos otra propiedad general de la transformación lineal T. Tal transformación «preserva» las combinaciones lineales; esto es, si $\alpha_1, \ldots, \alpha_n$ son vectores de V y $c_1, \ldots, c_n$ son escalares, entonces

$$T(c_1\alpha_1 + \cdots + c_n\alpha_n) = c_1(T\alpha_1) + \cdots + c_n(T\alpha_n).$$

Ello resulta directamente en forma en forma directa de la definición. Por ejemplo,

$$\begin{aligned} T(c_1\alpha_1 + c_2\alpha_2) &= c_1(T\alpha_1) + T(c_2\alpha_2) \\ &= c_1(T\alpha_1) + c_2(T\alpha_2). \end{aligned}$$

Teorema 1. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F, sea* $\{\alpha_1, \ldots, \alpha_n\}$ *una base ordenada de V. Sean W un espacio vectorial sobre el mismo cuerpo F y* $\beta_1, \ldots, \beta_n$ *vectores cualesquiera de W. Entonces existe una única transformación lineal de T de V en W tal que*

$$T\alpha_j = \beta_j, \qquad j = 1, \ldots, n.$$

Demostración. Para demostrar que existe una transformación lineal T tal que $T\alpha_j = \beta_j$ se procede como sigue. Dado α de V, existe una única n-tuple $(x_1, \ldots, x_n)$ tal que

$$\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n.$$

Para ese vector α se define

$$T\alpha = x_1\beta_1 + \cdots + x_n\beta_n.$$

Entonces, T es una correspondencia bien definida que asocia a cada vector α de V un vector $T\alpha$ de W. De la definición queda claro que $T\alpha_j = \beta_j$ para cada j. Para ver que T es lineal, sea

$$\beta = y_1\alpha_1 + \cdots + y_n\alpha_n$$

de V y sea c cualquier escalar. Ahora

$$c\alpha + \beta = (cx_1 + y_1)\alpha_1 + \cdots + (cx_n + y_n)\alpha_n$$

con lo que, por definición,

$$T(c\alpha + \beta) = (cx_1 + y_1)\beta_1 + \cdots + (cx_n + y_n)\beta_n.$$

Por otra parte,

$$\begin{aligned} c(T\alpha) + T\beta &= c\sum_{i=1}^{n} x_i\beta_i + \sum_{i=1}^{n} y_i\beta_i \\ &= \sum_{i=1}^{n} (cx_i + y_i)\beta_i \end{aligned}$$

y así

$$T(c\alpha + \beta) = c(T\alpha) + T\beta.$$

Si U es una transformación lineal de V en W con $U\alpha_j = \beta_j$, $j = 1, \ldots, n$, entonces para el vector $\alpha = \sum_{i=1}^{n} x_i\alpha_i$ se tiene

$$\begin{aligned} U\alpha &= U\left(\sum_{i=1}^{n} x_i\alpha_i\right) \\ &= \sum_{i=1}^{n} x_i(U\alpha_i) \\ &= \sum_{i=1}^{n} x_i\beta_i \end{aligned}$$

con lo que U es exactamente la misma correspondencia T que se definió antes, lo que demuestra que la transformación lineal T con $T\alpha_j = \beta_j$, es única. ∎

El Teorema 1 es muy elemental, pero por su importancia ha sido presentado detalladamente. El concepto de función es muy general. Si V y W son espacios vectoriales (no nulos), hay una multitud de funciones de V en W. El Teorema 1 destaca el hecho de que las funciones que son lineales son muy especiales.

Ejemplo 6. Los vectores

$$\alpha_1 = (1, 2)$$
$$\alpha_2 = (3, 4)$$

son linealmente independientes y, por tanto, forman una base de R^2. De acuerdo con el Teorema 1, existe una única transformación lineal de R^2 en R^2 tal que

$$T\alpha_1 = (3, 2, 1)$$
$$T\alpha_2 = (6, 5, 4)$$

De ser así, se debe poder encontrar $T(\epsilon_1)$. Encontrados los escalares c_1, c_2 tales que $\epsilon_1 = c_1\alpha_1 + c_2\alpha_2$, se sabe entonces que $T\epsilon_1 = c_1T\alpha_1 + c_2T\alpha_2$. Si $(1, 0) = c_1(1, 2) + c_2(3, 4)$, entonces $c_1 = -2$ y $c_2 = 1$. Con lo que

$$\begin{aligned} T(1, 0) &= -2(3, 2, 1) + (6, 5, 4) \\ &= (0, 1, 2). \end{aligned}$$

Ejemplo 7. Sea T una transformación lineal del espacio de los m-tuples F^m en el espacio de los n-tuples F^n. El Teorema 1 dice que T está unívocamente determinado por la sucesión de vectores $\beta_1, \ldots, \beta_m$, donde

$$\beta_i = T\epsilon_i, \qquad i = 1, \ldots, m.$$

Dicho brevemente, T está unívocamente determinado por las imágenes de los vectores de la base canónica. Esta determinación es

$$\alpha = (x_1, \ldots, x_m)$$
$$T\alpha = x_1\beta_1 + \cdots + x_m\beta_m.$$

Si B es la matriz $m \times n$ que tiene por filas los vectores $\beta_1, \ldots, \beta_m$, esto quiere decir que

$$T\alpha = \alpha B.$$

O sea que si $\beta_i = (B_{i1}, \ldots, B_{in})$, entonces

$$T(x_1, \ldots, x_n) = [x_1 \cdots x_m] \begin{bmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & & \vdots \\ B_{m1} & \cdots & B_{mn} \end{bmatrix}.$$

Esta es una descripción muy explícita de la transformación lineal. En la Sección 3.4 se hará un estudio detallado de la relación entre las transformaciones lineales y las matrices. No se seguirá con la descripción particular $T\alpha = \alpha B$, porque tiene la matriz B a la derecha del vector α y ello puede inducir a confusión. La razón de este ejemplo es hacer ver que se puede dar una descripción explícita y razonablemente simple de todas las transformaciones lineales de F^m en F^n.

Si T es una transformación lineal de V en W, entonces la imagen de T no es solo un subconjunto de W, sino un subespacio de W. Sea R_T la imagen de T; esto es, el conjunto de todos los vectores β de W tales que $\beta = T\alpha$, para algún α en V. Sean β_1 y β_2 de R_T y sea c un escalar. Existen vectores α_1 y α_2 de V tales que $T\alpha_1 = \beta_1$ y $T\alpha_2 = \beta_2$. Como T es lineal

$$\begin{aligned} T(c\alpha_1 + \alpha_2) &= cT\alpha_1 + T\alpha_2 \\ &= c\beta_1 + \beta_2, \end{aligned}$$

lo que dice que $c\beta_1 + \beta_2$ pertenece también a R_T.

Otro subespacio interesante asociado con la transformación lineal T es el conjunto N, que consta de los vectores α en V tales que $T\alpha = 0$: es un subespacio de V, pues

(a) $T(0) = 0$, con lo que N no es vacío.
(b) Si $T\alpha_1 = T\alpha_2 = 0$, entonces

$$\begin{aligned} T(c\alpha_1 + \alpha_2) &= cT\alpha_1 + T\alpha_2 \\ &= c0 + 0 \\ &= 0 \end{aligned}$$

con lo que $c\alpha_1 + \alpha_2$ pertenece a N.

Definición. *Sean V y W dos espacios vectoriales sobre el cuerpo F y sea T una transformación lineal de V en W. El* **espacio nulo** *de T es el conjunto de todos los vectores α de V tales que $T\alpha = 0$.*

Si V es de dimensión finita, el **rango** *de T es la dimensión de la imagen de T y la* **nulidad** *de T es la dimensión del espacio nulo de T.*

He aquí uno de los resultados más importantes del álgebra lineal.

Teorema 2. *Sean V y W espacios vectoriales sobre el cuerpo F y sea T una transformación lineal de V en W. Supóngase que V es de dimensión finita. Entonces*

$$\text{rango } (T) + \text{nulidad } (T) = \dim V.$$

Demostración. Sea $\{\alpha_1, \ldots, \alpha_k\}$ una base de N, el espacio nulo de T. Existen vectores $\alpha_{k+1}, \ldots, \alpha_n$ en V tales que $\{\alpha_1, \ldots, \alpha_n\}$ es una base de V. Podemos demostrar ahora que $\{T\alpha_{k+1}, \ldots, T\alpha_n\}$ es una base para la imagen de T. Los vectores $T\alpha_1, \ldots, T\alpha_n$ generan evidentemente la imagen de T, y como $T\alpha_j = 0$ para $j < k$, se ve que $T\alpha_{k+1}, \ldots, T\alpha_n$ generan la imagen. Para ver que estos vectores son linealmente independientes, supóngase que se tienen escalares c_i tales que

$$\sum_{i=k+1}^{n} c_i(T\alpha_i) = 0.$$

Esto dice que

$$T\left(\sum_{i=k+1}^{n} c_i\alpha_i\right) = 0$$

y en consecuencia, el vector $\alpha = \sum_{i=k+1}^{n} c_i\alpha_i$ pertenece al espacio nulo de T. Como $\alpha_1, \ldots, \alpha_k$ forman una base de N, deben existir escalares $b_1, \ldots, b_k$ tales que

$$\alpha = \sum_{i=1}^{k} b_i\alpha_i.$$

Con lo que

$$\sum_{i=1}^{k} b_i\alpha_i - \sum_{j=k+1}^{n} c_j\alpha_j = 0$$

y como $\alpha_1, \ldots, \alpha_n$ son linealmente independientes, se debe tener

$$b_1 = \cdots = b_k = c_{k+1} = \cdots = c_n = 0.$$

Si el rango de T es r, el hecho de que $T\alpha_{k+1}, \ldots, T\alpha_n$ formen una base de la imagen de T nos dice que $r = n - k$. Como k es la nulidad de T y n es la dimensión de V, se tiene lo afirmado. ∎

Teorema 3. *Si A es una matriz $m \times n$ de elementos en el cuerpo F, entonces*

$$\textit{rango de filas } (A) = \textit{rango de columnas } (A)$$

Demostración. Sea T una transformación lineal de $F^{n\times 1}$ en $F^{m\times 1}$ definida por $T(X) = AX$. El espacio nulo de T es el espacio de soluciones del sistema $AX = 0$, es decir, el conjunto de todas las matrices columna X tales que $AX = 0$. La imagen de T es el conjunto de todas las matrices columna $m \times 1$, Y, tales que $AX = Y$ tiene una solución para X. Si $A_1, \ldots, A_n$ son las columnas de A, entonces

$$AX = x_1A_1 + \cdots + x_nA_n$$

de modo que la imagen de T es el subespacio generado por las columnas de A. Es decir, que la imagen de T es el espacio de las columnas de A. Por tanto,

$$\text{rango } (T) = \text{rango de columnas } (A).$$

El Teorema 2 dice que si S es el espacio de soluciones del sistema $AX = 0$, entonces

$$\dim S + \text{rango de columnas } (A) = n.$$

Sea ahora el Ejemplo 15 del Capítulo 2. Ahí se vio que, si r es la dimensión del espacio de filas de A, entonces el espacio de soluciones S tiene una base que consta de $n - r$ vectores:

$$\dim S = n - \text{rango de filas } (A).$$

En consecuencia

$$\text{rango de filas } (A) = \text{rango de columnas } (A). \quad ∎$$

La demostración del Teorema 3 que acaba de darse depende de cálculos explícitos con sistemas de ecuaciones lineales. Hay una demostración más conceptual que no se basa en tales cálculos. Se dará tal demostración en la Sección 3.7.

Ejercicios

1. ¿Cuáles de las siguientes funciones T de R^2 en R^2 son transformaciones lineales?

(a) $T(x_1, x_2) = (1 + x_1, x_2)$;
(b) $T(x_1, x_2) = (x_2, x_1)$;
(c) $T(x_1, x_2) = (x_1^2, x_2)$;
(d) $T(x_1, x_2) = (\operatorname{sen} x_1, x_2)$;
(e) $T(x_1, x_2) = (x_1 - x_2, 0)$.

2. Hallar la imagen, rango, espacio nulo y nulidad para la transformación cero y la transformación identidad en un espacio de dimensión finita V.

3. Describir la imagen y el espacio nulo para la transformación derivación del Ejemplo 2. Hacer lo mismo para la transformación integración del Ejemplo 5.

4. ¿Existe una transformación lineal T de R^3 en R^2 tal que $T(1, -1, 1) = (1, 0)$ y $T(1, 1, 1) = (0, 1)$?

5. Si

$$\begin{aligned} \alpha_1 &= (1, -1), & \beta_1 &= (1, 0) \\ \alpha_2 &= (2, -1), & \beta_2 &= (0, 1) \\ \alpha_3 &= (-3, 2), & \beta_3 &= (1, 1) \end{aligned}$$

¿Existe una transformación lineal T de R^2 en R^2 tal que $T\alpha_i = \beta_i$ para $i = 1, 2, 3$?

6. Describir explícitamente (como en los Ejercicios 1 y 2) la transformación lineal T de F^2 en F^2 tal que $T\epsilon_1 = (a, b)$, $T\epsilon_2 = (c, d)$.

7. Sea F un subcuerpo de los números complejos y sea T una función de F^3 en F^3 definida por

$$T(x_1, x_2, x_3) = (x_1 - x_2 + 2x_3, 2x_1 + x_2, -x_1 - 2x_2 + 2x_3).$$

(a) Comprobar que T es una transformación lineal.
(b) Si (a, b, c) es un vector de F^3, ¿cuáles son las condiciones para a, b y c de modo que el vector pertenezca a la imagen de T? ¿Cuál es el rango de T?
(c) ¿Cuáles son las condiciones para a, b y c de modo que (a, b, c) pertenezca al espacio nulo de T? ¿Cuál es la nulidad de T?

8. Describir explícitamente una transformación lineal de R^3 en R^3 que tiene como imagen el subespacio generado por $(1, 0, -1)$ y $(1, 2, 2)$.

9. Sea V el espacio vectorial de todas las matrices $n \times n$ sobre el cuerpo F y sea B una matriz $n \times n$ dada. Si

$$T(A) = AB - BA$$

comprobar que T es una transformación lineal de V en V.

10. Sea V el conjunto de todos los números complejos considerado como un espacio vectorial sobre el cuerpo de los números *reales* (con las operaciones usuales). Hallar una función de V en V que sea una transformación lineal en dicho espacio vectorial, pero que no sea una transformación lineal en C^1; es decir, que no sea lineal compleja.

11. Sea V el espacio de las matrices $n \times 1$ sobre F y sea W el espacio de las matrices $m \times 1$ sobre F. Sea A una matriz $m \times n$ dada y sea T la transformación lineal de V en W, definida

por $T(X) = AX$. Demostrar que T es la transformación cero si, y solo si, A es la matriz cero.

12. Sea V un espacio vectorial V de dimensión n sobre el cuerpo F y sea T una transformación lineal de V en V tal que la imagen y el espacio nulo de T sean idénticos. Demostrar que n es par. (¿Se puede dar un ejemplo de tal transformación lineal T?)

13. Sea V un espacio vectorial y T una transformación lineal de V en V. Demostrar que las dos afirmaciones siguientes sobre T son equivalentes.

(a) La intersección de la imagen de T y el espacio nulo de T es el subespacio cero de V.
(b) Si $T(T\alpha) = 0$, entonces $T\alpha = 0$.

3.2. Algebra de las transformaciones lineales

En el estudio de las transformaciones lineales de V en W es de fundamental importancia que el conjunto de estas transformaciones hereda una estructura natural de espacio vectorial. El conjunto de las transformaciones lineales de un espacio V en sí mismo tiene incluso una estructura algebraica mayor, pues la composición ordinaria de funciones da una «multiplicación» de tales transformaciones. Se analizarán estas ideas en esta sección.

Teorema 4. *Sean V y W espacios vectoriales sobre el cuerpo F. Sean T y U transformaciones lineales de V en W. La función $(T + U)$ definida por*

$$(T + U)(\alpha) = T\alpha + U\alpha$$

es una transformación lineal de V en W. Si c es cualquier elemento de F, la función (cT) definida por

$$(cT)(\alpha) = c(T\alpha)$$

es una transformación lineal de V en W. El conjunto de todas las transformaciones lineales de V en W, junto con la adición y la multiplicación escalar aquí definidas, es un espacio vectorial sobre el cuerpo F.

Demostración. Supóngase que T y U son transformaciones lineales de V en W, que se define $(T + U)$ como se indicó. Entonces

$$\begin{aligned}(T + U)(c\alpha + \beta) &= T(c\alpha + \beta) + U(c\alpha + \beta)\\ &= c(T\alpha) + T\beta + c(U\alpha) + U\beta\\ &= c(T\alpha + U\alpha) + (T\beta + U\beta)\\ &= c(T + U)(\alpha) + (T + U)(\beta)\end{aligned}$$

que dice que $(T + U)$ es una transformación lineal. En forma análoga,

$$\begin{aligned}(cT)(d\alpha + \beta) &= c[T(d\alpha + \beta)]\\ &= c[d(T\alpha) + T\beta]\\ &= cd(T\alpha) + c(T\beta)\\ &= d[c(T\alpha)] + c(T\beta)\\ &= d[(cT)\alpha] + (cT)\beta\end{aligned}$$

que dice que (cT) es una transformación lineal.

Para comprobar que el conjunto de transformaciones lineales de V en W (junto con estas operaciones) es un espacio vectorial, se debe verificar directamente cada una de las condiciones para la adición vectorial y la multiplicación por escalar. Se dejan los detalles de esto al lector, bastando aquí los siguientes comentarios: El vector cero en este espacio será la transformación nula, que transforma cada vector de V en el vector cero de W; cada una de las propiedades de las dos operaciones es consecuencia de la correspondiente propiedad de las operaciones en el espacio W. ∎

Es conveniente mencionar otro modo de ver este teorema. Si se definen la suma y la multiplicación por escalar como se hizo antes, entonces el conjunto de todas las funciones de V en W es un espacio vectorial sobre el cuerpo F. Esto no tiene nada que ver con que V sea un espacio vectorial, solo necesita que V sea un conjunto no vacío. Cuando V es un espacio vectorial, se puede definir una transformación lineal de V en W, y el Teorema 4 dice que las transformaciones lineales forman un subespacio del espacio de todas las funciones de V en W.

Se representará el espacio de las transformaciones lineales de V en W por $L(V, W)$. Se recuerda al lector que $L(V, W)$ se define solo cuando V y W son espacios vectoriales sobre el mismo cuerpo.

Teorema 5. *Sea V un subespacio vectorial de dimensión finita n sobre el cuerpo F, y sea W un espacio vectorial de dimensión finita m sobre F. Entonces el espacio $L(V, W)$ es de dimensión finita y tiene dimensión mn.*

Demostración. Sean

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{y} \quad \mathcal{B}' = \{\beta_1, \ldots, \beta_m\}$$

bases ordenadas de V y W, respectivamente. Para cada par de enteros (p, q) con $1 \leq p \leq m$ y $1 \leq q \leq n$ se define una transformación lineal $E^{p,q}$ de V en W por

$$E^{p,q}(\alpha_i) = \begin{cases} 0, & \text{si } i \neq q \\ \beta_p, & \text{si } i = q \end{cases}$$

$$= \delta_{iq}\beta_p.$$

De acuerdo con el Teorema 1, existe una transformación lineal única de V en W que satisface estas condiciones. Se afirma que las mn transformaciones $E^{p,q}$ forman una base de $L(V, W)$.

Sea T una transformación lineal de V en W. Para cada j, $1 \leq j \leq n$, sean $A_{1j}, \ldots, A_{mj}$ las coordenadas del vector $T\alpha_j$ en la base ordenada $\mathcal{B}'$, es decir,

$$(3\text{-}1) \qquad T\alpha_j = \sum_{p=1}^{m} A_{pj}\beta_p.$$

Queremos demostrar que

$$(3\text{-}2) \qquad T = \sum_{p=1}^{m} \sum_{q=1}^{n} A_{pq}E^{p,q}$$

Sea U la transformación lineal del segundo miembro de (3-2). Entonces para cada j

$$\begin{aligned} U\alpha_j &= \sum_p \sum_q A_{pq}E^{p,q}(\alpha_j) \\ &= \sum_p \sum_q A_{pq}\delta_{jq}\beta_p \\ &= \sum_{p=1}^{m} A_{pj}\beta_p \\ &= T\alpha_j \end{aligned}$$

y en consecuencia $U = T$. Pero como (3-2) dice que los $E^{p,q}$ generan $L(V, W)$ debemos demostrar que son independientes. Pero esto queda claro con lo expuesto anteriormente; en efecto, si la transformación

$$U = \sum_p \sum_q A_{pq}E^{p,q}$$

es la transformación nula, entonces $U\alpha_j = 0$ para cada j, con lo que

$$\sum_{p=1}^{m} A_{pj}\beta_p = 0$$

y la independencia de los β_p implica que $A_{pj} = 0$ para todo p y j. ▮

Teorema 6. *Sean V, W y Z espacios vectoriales sobre el cuerpo F. Sea T una transformación lineal de V en W y U una transformación lineal de W en Z. Entonces la función compuesta UT definida por $UT(\alpha) = U(T(\alpha))$ es una transformación lineal de V en Z.*

Demostración.

$$\begin{aligned} (UT)(c\alpha + \beta) &= U[T(c\alpha + \beta)] \\ &= U(cT\alpha + T\beta) \\ &= c[U(T\alpha)] + U(T\beta) \\ &= c(UT)(\alpha) + (UT)(\beta). \end{aligned}$$ ▮

En lo que sigue debemos interesarnos principalmente en transformaciones lineales de un espacio vectorial en sí mismo. Como se tendrá a menudo que escribir «T es una transformación lineal de V en V», se dirá más bien: «T es un operador lineal sobre V».

Definición. *Si V es un espacio vectorial sobre el cuerpo F, un* **operador lineal** *sobre V es una transformación lineal de V en V.*

En el caso del Teorema 6, cuando $V = W = Z$, en que U y T son operadores lineales en el espacio V, se ve que la composición UT es también un operador lineal sobre V. Así, el espacio $L(V, V)$ tiene una «multiplicación» definida por composición. En este caso el operador TU también está definido, y debe observarse que en general $UT \neq TU$, es decir, $UT - TU \neq 0$. Se ha de advertir de manera especial que si T es un operador lineal sobre V, entonces

se puede componer T con T. Se usará para ello la notación $T^2 = TT$, y en general $T^n = T \cdots T$ (n veces) para $n = 1, 2, 3, \ldots$ Se define $T^0 = I$ si $T \neq 0$.

Lema. *Sea V un espacio vectorial sobre el cuerpo F; sean U, T_1 y T_2 operadores lineales sobre V; sea c un elemento de F.*

(a) $IU = UI = U$;
(b) $U(T_1 + T_2) = UT_1 + UT_2$; $(T_1 + T_2)U = T_1U + T_2U$;
(c) $c(UT_1) = (cU)T_1 = U(cT_1)$.

Demostración. (a) Esta propiedad de la función identidad es obvia. Se la ha enunciado aquí solo para insistir.

(b)
$$\begin{aligned}[U(T_1 + T_2)](\alpha) &= U[(T_1 + T_2)(\alpha)] \\ &= U(T_1\alpha + T_2\alpha) \\ &= U(T_1\alpha) + U(T_2\alpha) \\ &= (UT_1)(\alpha) + (UT_2)(\alpha)\end{aligned}$$

así $U(T_1 + T_2) = UT_1 + UT_2$. También
$$\begin{aligned}[(T_1 + T_2)U](\alpha) &= (T_1 + T_2)(U\alpha) \\ &= T_1(U\alpha) + T_2(U\alpha) \\ &= (T_1U)(\alpha) + (T_2U)(\alpha)\end{aligned}$$

con lo que $(T_1 + T_2)U = T_1U + T_2U$. (El lector observará que las demostraciones de estas dos leyes distributivas no han tenido en cuenta que T_1 y T_2 eran lineales, y la demostración de la segunda tampoco considera el que U es lineal.)

(c) Se deja al lector la demostración de la parte (c). ▮

El contenido de este lema, y una parte del Teorema 5, dicen que el espacio vectorial $L(V, V)$, junto con la operación de composición, es lo que se conoce como un álgebra lineal con identidad. Se examinará esto en el Capítulo 4.

Ejemplo 8. Si A es una matriz $m \times n$ con elementos en F, se tiene la transformación lineal T definida por $T(X) = AX$ de $F^{n \times 1}$ en $F^{m \times 1}$. Si B es una matriz $p \times m$, se tiene la transformación lineal U de $F^{m \times 1}$ en $F^{p \times 1}$ definida por $U(Y) = BY$. La composición UT se define fácilmente por:

$$\begin{aligned}(UT)(X) &= U(T(X)) \\ &= U(AX) \\ &= B(AX) \\ &= (BA)X.\end{aligned}$$

Así UT es «multiplicación a la izquierda por el producto de matrices BA».

Ejemplo 9. Sea F un cuerpo y V el espacio vectorial de todas las funciones polinomios de F en F. Sea D el operador de derivación definido en el Ejemplo 2, y sea T el operador lineal «multiplicación por x»:

$$(Tf)(x) = xf(x).$$

Entonces $DT \neq TD$. En efecto, el lector no tendrá dificultad en verificar que $DT - TD = I$, el operador identidad.

Aun cuando la «multiplicación» que se tiene en $L(V, V)$ no es conmutativa, está muy relacionada con las operaciones en el espacio vectorial $L(V, V)$.

Ejemplo 10. Sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ordenada de un espacio vectorial V. Se consideran los operadores lineales $E^{p,q}$ que se presentaron en la demostración del Teorema 5:

$$E^{p,q}(\alpha_i) = \delta_{iq}\alpha_p.$$

Estos n^2 operadores lineales forman una base del espacio de los operadores lineales sobre V. ¿Qué es $E^{p,q}E^{r,s}$? Se tiene

$$\begin{aligned}(E^{p,q}E^{r,s})(\alpha_i) &= E^{p,q}(\delta_{is}\alpha_r)\\ &= \delta_{is}E^{p,q}(\alpha_r)\\ &= \delta_{is}\delta_{rq}\alpha_p.\end{aligned}$$

Luego

$$E^{p,q}E^{r,s} = \begin{cases}0, & \text{si} \quad r \neq q\\ E^{p,s}, & \text{si} \quad q = r.\end{cases}$$

Sea T un operador lineal sobre V. Se probó en la demostración del Teorema 5 que si

$$\begin{aligned}A_j &= [T\alpha_j]_{\mathcal{B}}\\ A &= [A_1, \ldots, A_n]\end{aligned}$$

entonces

$$T = \sum_p \sum_q A_{pq}E^{p,q}.$$

Si

$$U = \sum_r \sum_s B_{rs}E^{r,s}$$

es otro operador lineal en V, entonces el lema anterior dice que

$$\begin{aligned}TU &= (\sum_p \sum_q A_{pq}E^{p,q})(\sum_r \sum_s B_{rs}E^{r,s})\\ &= \sum_p \sum_q \sum_r \sum_s A_{pq}B_{rs}E^{p,q}E^{r,s}.\end{aligned}$$

Como observamos, los únicos términos que quedan en esta considerable suma son los términos con $q = r$, y como $E^{p,r}E^{r,s} = E^{p,s}$, se tiene

$$\begin{aligned}TU &= \sum_p \sum_s (\sum_r A_{pr}B_{rs})E^{p,s}\\ &= \sum_p \sum_s (AB)_{ps}E^{p,s}.\end{aligned}$$

Así el efecto de componer T y U es el de multiplicar las matrices A y B.

En el tratamiento de las operaciones algebraicas con transformaciones lineales no se ha dicho nada aún sobre inversión. Una pregunta concreta de interés es la siguiente. ¿Para cuáles operadores lineales T en el espacio V existe un operador lineal T^{-1} tal que $TT^{-1} = T^{-1}T = I$?

Una función T de V en W se dice **inversible** si existe una función U de W en V tal que UT es la función identidad de V y TU es la función identidad de W. Si T es inversible, la función U es única y se representa por T^{-1}. (Véase Apéndice.) Más aún, T es inversible si y, solo si,

1. T es inyectiva, esto es, $T\alpha = T\beta$ implica $\alpha = \beta$.
2. T es sobreyectiva, esto es, la imagen de T es (coincide con) W.

Teorema 7. *Sean V y W dos espacios vectoriales sobre el cuerpo F y sea T una transformación lineal de V en W. Si T es inversible, entonces la función recíproca T^{-1} es una transformación lineal de W sobre V.*

Demostración. Volvemos a repetir para aclarar un aspecto. Cuando T es una función inyectiva y sobreyectiva, existe una única función recíproca T^{-1} que aplica W sobre V, de modo que $T^{-1}T$ sea la función identidad de V, y TT^{-1} sea la función identidad de W. Lo que se ha de demostrar aquí es que si una función lineal T es inversible, entonces la recíproca T^{-1} también es lineal.

Sean β_1 y β_2 dos vectores de W y sea c un escalar. Queremos demostrar que

$$T^{-1}(c\beta_1 + \beta_2) = cT^{-1}\beta_1 + T^{-1}\beta_2.$$

Sea $\alpha_i = T^{-1}\beta_i$, $i = 1, 2$; esto es, sea α_i el único vector de V, tal que $T\alpha_i = \beta_i$. Como T es lineal,

$$\begin{aligned} T(c\alpha_1 + \alpha_2) &= cT\alpha_1 + T\alpha_2 \\ &= c\beta_1 + \beta_2. \end{aligned}$$

Así, $c\alpha_1 + \alpha_2$ es el único vector de V que es aplicado por T en $c\beta_1 + \beta_2$, y así

$$\begin{aligned} T^{-1}(c\beta_1 + \beta_2) &= c\alpha_1 + \alpha_2 \\ &= c(T^{-1}\beta_1) + T^{-1}\beta_2 \end{aligned}$$

y T^{-1} es lineal. ∎

Supóngase que se tiene una transformación lineal inversible T de V sobre W y una transformación lineal inversible U de W sobre Z. Entonces UT es inversible y $(UT)^{-1} = T^{-1}U^{-1}$. Esta conclusión no exige la linealidad, ni tampoco implica comprobar separadamente que UT es inyectiva y sobreyectiva. Todo lo que se necesita es comprobar que $T^{-1}U^{-1}$ es inversa a la izquierda e inversa a la derecha de UT.

Si T es lineal, entonces $T(\alpha - \beta) = T\alpha - T\beta$; luego $T\alpha = T\beta$ si y, solo si, $T(\alpha - \beta) = 0$. Esto simplifica mucho la comprobación de que T es inyectiva. Se dice que la transformación lineal T es **no singular** si $T\gamma = 0$ implica $\gamma = 0$; es decir, si el espacio nulo de T es $\{0\}$. Evidentemente, T es inyectiva si y, solo si, T es no singular. El alcance de esta observación es que las transformaciones lineales no singulares son las que preservan la independencia lineal.

Teorema 8. *Sea T una transformación lineal de V en W. Entonces T es no singular si, y solo si, T aplica cada subconjunto linealmente independiente de V sobre un subconjunto linealmente independiente de W.*

Demostración. Supóngase primero que T es no singular. Sea S un subconjunto linealmente independiente de V. Si $\alpha_1, \ldots, \alpha_k$ son vectores pertenecientes a S, entonces los vectores $T\alpha_1, \ldots, T\alpha_k$ son linealmente independientes; en efecto, si

$$c_1(T\alpha_1) + \cdots + c_k(T\alpha_k) = 0$$

entonces

$$T(c_1\alpha_1 + \cdots + c_k\alpha_k) = 0$$

y como T es no singular

$$c_1\alpha_1 + \cdots + c_k\alpha_k = 0$$

de lo que se sigue con que cada $c_i = 0$, pues S es un conjunto independiente. Este razonamiento muestra que la imagen de S por T es independiente.

Supóngase que T aplica subconjuntos independientes sobre subconjuntos independientes. Sea α un vector no nulo de V. Entonces el conjunto S que consta del solo vector α es independiente. La imagen de S es el conjunto que consta del solo vector $T\alpha$. Por tanto, $T\alpha \neq 0$, pues el conjunto que consta del solo vector nulo es dependiente; lo que muestra que el espacio nulo de T es el subespacio cero, es decir, T es no singular.

Ejemplo 11. Sea F un subcuerpo de los números complejos (o un cuerpo de característica cero) y sea V el espacio de las funciones polinomios sobre F. Considérese el operador derivación D y el operador de la «multiplicación por x», T, del Ejemplo 9. Como D aplica todas las constantes sobre 0, D es singular; sin embargo, V no es de dimensión finita, la imagen de D es todo V y es posible definir una inversa a la derecha de D. Por ejemplo, si E es el operador integración indefinida:

$$E(c_0 + c_1x + \cdots + c_nx^n) = c_0x + \frac{1}{2}c_1x^2 + \cdots + \frac{1}{n+1}c_nx^{n+1}$$

entonces E es un operador lineal en V y $DE = I$. Por otro lado, $ED \neq I$, pues ED aplica las constantes sobre 0. El operador T está, por así decirlo, en la situación contraria. Si $xf(x) = 0$ para todo x, entonces $f = 0$. Con lo que T es no singular y es posible hallar una inversa a la izquierda de T. Por ejemplo, si U es la operación «suprimir el término constante y dividir por x»:

$$U(c_0 + c_1x + \cdots + c_nx^n) = c_1 + c_2x + \cdots + c_nx^{n-1}$$

entonces U es un operador lineal en V y $UT = I$. Pero $TU \neq I$, ya que cada función en la imagen de TU está en la imagen de T, que es el espacio de las funciones polinomios f tales que $f(0) = 0$.

Ejemplo 12. Sea F un cuerpo y sea T el operador lineal sobre F^2 definido por

$$T(x_1, x_2) = (x_1 + x_2, x_1).$$

Entonces T es no singular, pues si $T(x_1, x_2) = 0$ se tiene

$$\begin{aligned} x_1 + x_2 &= 0 \\ x_1 &= 0 \end{aligned}$$

con lo que $x_1 = x_2 = 0$. También se ve que T es sobreyectiva, pues si (z_1, z_2) es cualquier vector de F^2, para ver que (z_1, z_2) pertenece a la imagen de T se han de encontrar escalares x_1 y x_2 tales que

$$\begin{aligned} x_1 + x_2 &= z_1 \\ x_1 &= z_2 \end{aligned}$$

y la solución obvia es $x_1 = z_2$, $x_2 = z_1 - z_2$. Este último cálculo da una fórmula explícita para T^{-1}, a saber,

$$T^{-1}(z_1, z_2) = (z_2, z_1 - z_2).$$

Se ha visto en el Ejemplo 11 que una transformación lineal puede ser no singular sin ser sobreyectiva y que puede ser sobreyectiva sin ser no singular. El presente ejemplo ilustra un importante caso en que ello no puede suceder.

Teorema 9. *Sean V y W espacios vectoriales de dimensión finita sobre el cuerpo F tal que $\dim V = \dim W$. Si T es una transformación lineal de V en W, las siguientes afirmaciones son equivalentes:*

(i) *T es inversible.*
(ii) *T es no singular.*
(iii) *T es sobreyectiva; eso es, la imagen de T es W.*

Demostración. Sea $n = \dim V = \dim W$. Por el Teorema 2 se sabe que

$$\text{rango de } (T) + \text{nulidad } (T) = n.$$

Ahora bien, T es no singular si, y solo si, nulidad $(T) = 0$, y (como $n = \dim W$) la imagen de T es W si, y solo si, rango $(T) = n$. Como el rango más la nulidad es n, la nulidad es 0 precisamente cuando el rango es n. Por tanto, T es no singular si, y solo si, $T(V) = W$. Así, si rigen las condiciones (ii) o (iii), la otra se cumple también y T es inversible. ∎

Se previene al lector que no debe aplicar el Teorema 9, excepto si la dimensión es finita y con $\dim V = \dim W$. Con las hipótesis del Teorema 9 las condiciones (i), (ii) y (iii) son también equivalentes a éstas.

(iv) *Si $\{\alpha_1, \ldots, \alpha_n\}$ es una base de V, entonces $\{T\alpha_1, \ldots, T\alpha_n\}$ es una base de W.*

(v) *Existe una base $\{\alpha_1, \ldots, \alpha_n\}$ de V tal que $\{T\alpha_1, \ldots, T\alpha_n\}$ es una base de W.*

Se dará una demostración de la equivalencia de las cinco condiciones que es diferente a la dada para la equivalencia de (i), (ii) y (iii).

(i) $\rightarrow$ (ii). Si T es inversible, T es no singular. (ii) $\rightarrow$ (iii). Supóngase que T es no singular. Sea $\{\alpha_1, \ldots, \alpha_n\}$ una base de V. Por el Teorema 8, $\{T\alpha_1, \ldots, T\alpha_n\}$ es un conjunto linealmente independiente de vectores de W, y como la

dimensión de W es también n, este conjunto de vectores es una base de W. Ahora sea β cualquier vector de W. Existen escalares $c_1, \ldots, c_n$ tales que

$$\begin{aligned}\beta &= c_1(T\alpha_1) + \cdots + c_n(T\alpha_n)\\ &= T(c_1\alpha_1 + \cdots + c_n\alpha_n)\end{aligned}$$

lo que muestra que β pertenece a la imagen de T. (iii) $\longrightarrow$ (iv). Se supone ahora que T es sobreyectiva. Si $\{\alpha_1, \ldots, \alpha_n\}$ es cualquier base de V, los vectores $T\alpha_1, \ldots, T\alpha_n$ generan la imagen de T, que es W por definición. Como la dimensión de W es n, estos n vectores deben ser linealmente independientes, esto es, deben constituir una base de W. (iv) $\longrightarrow$ (v). Esto no requiere comentarios. (v) $\longrightarrow$ (i). Supóngase que existe alguna base $\{\alpha_1, \ldots, \alpha_n\}$ de V tal que $\{\alpha_1, \ldots, T\alpha_n\}$ es una base de W. Como los $T\alpha_i$ generan W, está claro que la imagen de T es todo W. Si $\alpha = c_1\alpha_1 + \cdots + c_n\alpha_n$ pertenece al espacio nulo de T, entonces

$$T(c_1\alpha_1 + \cdots + c_n\alpha_n) = 0$$

o

$$c_1(T\alpha_1) + \cdots + c_n(T\alpha_n) = 0$$

y como los $T\alpha_i$ son independientes, cada $c_i = 0$, y así $\alpha = 0$. Se ha visto que el recorrido de T es W, y que T es no singular, luego T es inversible.

El conjunto de operadores lineales inversibles sobre un espacio V con la operación de composición proporciona un buen ejemplo de lo que se conoce en álgebra como «grupo». Aunque no se tendrá tiempo para examinar los grupos con algún detalle, se dará al menos la definición.

Definición. *Un* **grupo** *consta de lo siguiente.*

1. *Un conjunto G;*
2. *Una correspondencia (u operación) que asocia a cada par de elementos x, y de G, un elemento xy de G de tal modo que*
   (a) *$x(yz) = (xy)z$ para todo x, y, z en G (asociatividad);*
   (b) *existe un elemento e en G tal que $ex = xe = x$ para todo x de G;*
   (c) *a cada elemento x de G le corresponde un elemento x^{-1} en G tal que $xx^{-1} = x^{-1}x = e$.*

Se ha visto que la composición $(U, T) \longrightarrow UT$ asocia a cada par de operadores lineales inversibles sobre un espacio V otro operador inversible sobre V. La composición es una operación asociativa. El operador identidad I satisface $IT = TI = T$ para todo T, y para un T inversible existe (por el Teorema 7) un operador lineal inversible T^{-1} tal que $TT^{-1} = T^{-1}T = I$. Con lo que el conjunto de los operadores lineales inversibles sobre V, junto con esta operación, es un grupo. El conjunto de las matrices $n \times n$ inversibles, con la multiplicación matricial como operación, es otro ejemplo de grupo. Un grupo se dice **conmutativo** si satisface la condición $xy = yx$ para cada x e y. Los dos ejemplos dados anteriormente no son, en general, grupos conmutativos. A menudo se escribe la operación en un grupo conmutativo como $(x, y) \longrightarrow x + y$, en vez de $(x, y) \longrightarrow xy$, usándose entonces el símbolo 0 para el elemento «neutro» e.

El conjunto de los vectores en un espacio vectorial, junto con la operación de adición vectorial, es un grupo conmutativo. Un cuerpo puede ser descrito como un conjunto con dos operaciones, llamadas adición y multiplicación, que es un grupo conmutativo para la adición, y en que los elementos no nulos forman un grupo conmutativo para la multiplicación, teniéndose, además, la ley distributiva $x(y + z) = xy + xz$.

Ejercicios

1. Sean T y U operaciones lineales en R^2 definidas por

$$T(x_1, x_2) = (x_2, x_1) \quad \text{y} \quad U(x_1, x_2) = (x_1, 0).$$

(a) ¿Cómo se describirán T y U geométricamente?

(b) Dar reglas semejantes a las que definieron T y U para cada una de las transformaciones $(U + T)$, UT, TU, T^2, U^2.

2. Sea T el (único) operador lineal sobre C^3 para el que

$$T\epsilon_1 = (1, 0, i), \qquad T\epsilon_2 = (0, 1, 1), \qquad T\epsilon_3 = (i, 1, 0).$$

¿Es T inversible?

3. Sea T el operador lineal sobre R^3 definido por

$$T(x_1, x_2, x_3) = (3x_1, x_1 - x_2, 2x_1 + x_2 + x_3).$$

¿Es T inversible? De serlo, hallar una expresión para T^{-1} como aquella que define a T.

4. Para el operador lineal T del Ejercicio 3, demostrar que

$$(T^2 - I)(T - 3I) = 0.$$

5. Sea $C^{2\times 2}$ el espacio vectoriál complejo de las matrices 2×2 de elementos complejos. Sea

$$B = \begin{bmatrix} 1 & -1 \\ -4 & 4 \end{bmatrix}$$

y sea T el operador lineal sobre $C^{2\times 2}$ definido por $T(A) = BA$. ¿Cuál es el rango de T? ¿Se puede describir T^2?

6. Sea T una transformación lineal de R^3 en R^2 y sea U una transformación lineal de R^2 en R^3. Demostrar que la transformación UT no es inversible. Generalizar el teorema.

7. Encontrar dos operadores lineales T y U en R^2 tales que $TU = 0$, pero que $UT \neq 0$.

8. Sean V un espacio vectorial sobre el cuerpo F y T un operador lineal sobre V. Si $T^2 = 0$, ¿qué se puede decir respecto a la relación entre la imagen de T y el espacio nulo de T? Dar un ejemplo de un operador lineal T en R^2 tal que $T^2 = 0$, pero $T \neq 0$.

9. Sea T un operador lineal sobre el espacio V de dimensión finita. Supóngase que existe un operador lineal U en V tal que $TU = I$. Demostrar que T es inversible y que $U = T^{-1}$. Dar un ejemplo que muestre que esto es falso cuando V no es de dimensión finita. (*Indicación*: Sea $T = D$ el operador derivación en el espacio de las funciones polinomios.)

10. Sea A una matriz $m \times n$ con elementos en F y sea T la transformación lineal de $F^{n\times 1}$

en $F^{m\times 1}$ definida por $T(X) = AX$. Hacer ver que, si $m < n$, puede suceder que T sea so breyectiva sin ser no singular. En forma semejante hacer ver que si $m > n$ puede tomars T no singular, pero no sobreyectiva.

11. Sea V un espacio vectorial de dimensión finita y sea T un operador lineal sobre V Supóngase que el rango (T^2) = rango (T). Demostrar que la imagen y el espacio nulo d T son disjuntos, es decir, tienen solo el vector nulo en común.

12. Sean p, m y n enteros positivos y F un cuerpo. Sean V el espacio de las matrices $m \times$ sobre F y W el espacio de las matrices $p \times n$ sobre F. Sea B una matriz $p \times m$ dada y se T la transformación lineal de V en W definida por $T(A) = BA$. Demostrar que T es inver sible si, y solo si, $p = m$ y B es una matriz $m \times m$ inversible.

3.3. *Isomorfismo*

Si V y W son espacios vectoriales sobre el cuerpo F, toda transformació lineal T de V en W sobreyectiva e inyectiva, se dice **isomorfismo de V sobre W** Si existe un isomorfismo de V sobre W, se dice que V es **isomorfo** a W.

Obsérvese que V es trivialmente isomorfo a V, ya que el operador identida es un isomorfismo de V sobre V. También si V es isomorfo a W por un isomor fismo T, entonces W es isomorfo a V, pues T^{-1} es un isomorfismo de W sobr V. El lector podrá demostrar fácilmente que si V es isomorfo a W y W es iso morfo a Z, entonces V es isomorfo a Z. En resumen, el isomorfismo es un relación de equivalencia sobre la clase de espacios vectoriales. Si existe un iso morfismo de V sobre W, se dirá a veces que V y W son isomorfos, en vez d que V es isomorfo a W. Ello no será motivo de confusión porque V es isomorf a W, si, y solo si, W es isomorfo a V.

Teorema 10. *Todo espacio vectorial de dimensión n sobre el cuerpo F e isomorfo al espacio F^n.*

Demostración. Sea V un espacio vectorial de dimensión n sobre el cuer po F y sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ordenada para V. Se define una función de V en F^n, como sigue: Si α pertenece a V, sea $T\alpha$ el n-tuple $(x_1, \ldots, x_n)$ d coordenadas de α respecto de la base ordenada $\mathcal{B}$, es decir, el n-tuple tal qu

$$\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n.$$

En nuestro estudio sobre coordenadas en el Capítulo 2, se verificó que T e lineal inyectiva y aplica V sobre F^n. ▮

Para muchos fines es frecuente considerar los espacios vectoriales isomor fos como si fueran «el mismo», aunque los vectores y las operaciones en lo espacios sean muy diferentes; es decir, a menudo se identifican espacios iso morfos. No pretendemos prolongar ahora la discusión de esta idea, sino qu la asimilación de la comprensión del isomorfismo y del sentido en que los es pacios isomorfos son «los mismos» se irá haciendo a medida que se continú el estudio de espacios vectoriales.

Hacemos algunos breves comentarios. Supóngase que T es un isomorfismo de V sobre W. Si S es un subconjunto de V, entonces el Teorema 8 dice que S es linealmente independiente si, y solo si, el conjunto $T(S)$ en W es independiente. Así, para decidir si S es independiente no importa si se considera S o $T(S)$. Por lo que se ve que un isomorfismo «preserva la dimensión», esto es, todo subespacio de dimensión finita de V tiene la misma dimensión que su imagen por T. He aquí una sencilla ilustración de esta idea. Supóngase que A es una matriz $m \times n$ sobre el cuerpo F. Ya se han dado en realidad dos definiciones del espacio solución de la matriz A. La primera es el conjunto de todos los n-tuples $(x_1, \ldots, x_n)$ de F^n que satisfacen a cada una de las ecuaciones del sistema $AX = 0$. La segunda es el conjunto de todas las matrices columnas $n \times 1$, X, tales que $AX = 0$. El primer espacio solución es así un subespacio F^n y el segundo es un subespacio del espacio de todas las matrices $n \times 1$ sobre F. Ahora bien, existe un isomorfismo completamente obvio entre F^n y $F^{n \times 1}$, a saber,

$$(x_1, \ldots, x_n) \rightarrow \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Por este isomorfismo, el primer espacio solución de A es aplicado sobre el segundo espacio solución. Estos espacios tienen la misma dimensión, y así, si se quiere demostrar un teorema respecto a la dimensión del espacio solución, no importa qué espacio se elija para analizar. En realidad, el lector no se sorprenderá si decidimos identificar F^n con el espacio de las matrices $n \times 1$. Esto se hará cuando sea conveniente, y cuando no sea conveniente no se hará.

Ejercicios

1. Sea V el conjunto de los números complejos y sea F el cuerpo de los números reales. Con las operaciones usuales V es un espacio vectorial sobre F. Describir explícitamente un isomorfismo de este espacio sobre R^2.

2. Sea V un espacio vectorial sobre el cuerpo de los números complejos y supóngase que existe un isomorfismo T de V sobre C^3. Sean α_1, α_2, α_3, α_4 vectores en V tales que

$$T\alpha_1 = (1, 0, i), \qquad T\alpha_2 = (-2, 1 + i, 0),$$
$$T\alpha_3 = (-1, 1, 1), \qquad T\alpha_4 = (\sqrt{2}, i, 3).$$

(a) ¿Está α_1 en el subespacio generado por α_2 y α_3?

(b) Sea W_1 el subespacio generado por α_1 y α_2 y sea W_2 el subespacio generado por α_3 y α_4. ¿Cuál es la intersección de W_1 y W_2?

(c) Hallar una base del subespacio de V generado por los cuatro vectores α_i.

3. Sea W el conjunto de todas las matrices hermíticas 2×2, esto es, el conjunto de las matrices complejas 2×2, A, tales que $A_{ij} = \bar{A}_{ji}$ (la barra indica conjugación compleja). Como se señalaba en el Ejemplo 6 del Capítulo 2, W es un espacio vectorial sobre el cuerpo de los números *reales* con las operaciones usuales. Verificar que

$$(x, y, z, t) \rightarrow \begin{bmatrix} t + x & y + iz \\ y - iz & t - x \end{bmatrix}$$

es un isomorfismo de R^4 sobre W.

4. Demostrar que $F^{m \times n}$ es isomorfo a F^{mn}.

5. Sea V el conjunto de los números complejos considerados como espacio vectorial sobre el cuerpo de los números reales (Ejercicio 1). Se define como sigue una función T de V en el espacio de las matrices reales 2×2. Si $z = x + iy$, con x e y números reales, entonces

$$T(z) = \begin{bmatrix} x + 7y & 5y \\ -10y & x - 7y \end{bmatrix}.$$

(a) Verificar que T es una transformación lineal (real) inyectiva de V en el espacio de las matrices reales 2×2.
(b) Verificar que $T(z_1 z_2) = T(z_1)T(z_2)$.
(c) ¿Cómo se describirá la imagen de T?

6. Sean V y W espacios vectoriales de dimensión finita sobre el cuerpo F. Demostrar que V y W son isomorfos si, y solo si, $\dim V = \dim W$.

7. Sean V y W espacios vectoriales sobre el cuerpo F y sea U un isomorfismo de V sobre W. Demostrar que $T \to UTU^{-1}$ es un isomorfismo de $L(V, V)$ sobre $L(W, W)$.

3.4. Representación de transformaciones por matrices

Sea V un espacio vectorial de dimensión n sobre el cuerpo F, y sea W un espacio vectorial de dimensión m sobre F. Sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ordenada de V, y $\mathcal{B}' = \{\beta_1, \ldots, \beta_m\}$ una base ordenada de W. Si T es cualquier transformación lineal de V en W, entonces T está determinada por su efecto sobre los vectores α_j. Cada uno de los n vectores $T\alpha_j$ se expresa de manera única como combinación lineal

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i \tag{3-3}$$

de los β_i, los escalares $A_{1j}, \ldots, A_{mj}$ son las coordenadas de $T\alpha_j$ en la base ordenada $\mathcal{B}'$. Por consiguiente, la transformación T está determinada por los mn escalares A_{ij} mediante la expresión (3-3). La matriz $m \times n$, A, definida por $A(i, j) = A_{ij}$ se llama **matriz de T respecto al par de bases ordenadas $\mathcal{B}$ y $\mathcal{B}'$**. La tarea inmediata es comprender claramente cómo la matriz A determina la transformación lineal T.

Si $\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n$ es un vector de V, entonces

$$\begin{aligned} T\alpha &= T\left(\sum_{j=1}^{n} x_j\alpha_j\right) \\ &= \sum_{j=1}^{n} x_j(T\alpha_j) \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\beta_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j\right) \beta_i. \end{aligned}$$

Si X es la matriz de las coordenadas de α en la base ordenada $\mathcal{B}$, entonces el cálculo anterior muestra que AX es la matriz de las coordenadas del vector $T\alpha$ en la base ordenada $\mathcal{B}'$, ya que el escalar

$$\sum_{j=1}^{n} A_{ij}x_j$$

es el elemento de la i-ésima fila de la matriz columna AX. Obsérvese también que si A es cualquier matriz $m \times n$ sobre el cuerpo F, entonces

$$T\left(\sum_{j=1}^{n} x_j\alpha_j\right) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}x_j\right)\beta_i \tag{3-4}$$

define una transformación lineal T de V en W, la matriz de la cual es A, respecto a $\mathcal{B}$, $\mathcal{B}'$. Resumiendo formalmente se tiene:

Teorema 11. *Sean V un espacio vectorial de dimensión n sobre el cuerpo F, y W un espacio vectorial de dimensión m sobre F. Sean $\mathcal{B}$ una base ordenada de V y $\mathcal{B}'$ una base ordenada de W. Para cada transformación lineal T de V en W, existe una matriz $m \times n$, A, cuyos elementos pertenecen a F, tal que*

$$[T\alpha]_{\mathcal{B}'} = A[\alpha]_{\mathcal{B}}$$

para todo vector α en V. Además, $T \rightarrow A$ es una correspondencia biyectiva entre el conjunto de todas las transformaciones lineales de V en W y el conjunto de todas las matrices $m \times n$ sobre el cuerpo F.

La matriz A, que está asociada a T en el Teorema 11, se llama la **matriz de T respecto a las bases ordenadas** $\mathcal{B}$, $\mathcal{B}'$. Obsérvese que la ecuación (3-3) dice que A es la matriz cuyas columnas $A_1, \ldots, A_n$ son dadas por

$$A_j = [T\alpha_j]_{\mathcal{B}'}, \qquad j = 1, \ldots, n.$$

Si U es otra transformación lineal de V en W y $B = [B_1, \ldots, B_n]$ es la matriz de U respecto a las bases ordenadas $\mathcal{B}$, $\mathcal{B}'$, entonces $cA + B$ es la matriz de $cT + U$ respecto a $\mathcal{B}$, $\mathcal{B}'$. Esto es claro porque

$$\begin{aligned} cA_j + B_j &= c[T\alpha_j]_{\mathcal{B}'} + [U\alpha_j]_{\mathcal{B}'} \\ &= [cT\alpha_j + U\alpha_j]_{\mathcal{B}'} \\ &= [(cT + U)\alpha_j]_{\mathcal{B}'}. \end{aligned}$$

Teorema 12. *Sea V un espacio vectorial de dimensión n sobre el cuerpo F. Para cada par de bases ordenadas $\mathcal{B}$, $\mathcal{B}'$ de V y W, respectivamente, la función que asigna a una transformación lineal T su matriz respecto a $\mathcal{B}$, $\mathcal{B}'$ es un isomorfismo entre el espacio $L(V, W)$ y el espacio de todas las matrices $m \times n$ sobre el cuerpo F.*

Demostración. Se observó antes que tal función es lineal y, como se estableció en el Teorema 11, esta función es inyectiva y aplica $L(V, W)$ sobre el conjunto de matrices $m \times n$. ∎

Estamos particularmente interesados en la representación por matrices de las transformaciones lineales de un espacio en sí mismo, es decir, de los operadores lineales sobre un espacio V. En tal caso es más conveniente usar la misma

base ordenada, esto es, hacer $\mathcal{B} = \mathcal{B}'$, y se dirá simplemente que la matriz que la representa es la **matriz de T respecto a la base ordenada** $\mathcal{B}$. Como este concepto será muy importante para este estudio, repasaremos su definición. Si T es un operador lineal sobre un espacio vectorial V de dimensión finita, y $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ es una base ordenada de V, la matriz de T respecto a $\mathcal{B}$ (o la matriz de T en la base ordenada $\mathcal{B}$) es la matriz $n \times n$, A, cuyos elementos A_{ij} están definidos por las ecuaciones

(3-5) $$T\alpha_j = \sum_{i=1}^{n} A_{ij}\alpha_i, \qquad j = 1, \ldots, n.$$

Se debe recordar siempre que esta matriz que representa a T depende de la base ordenada $\mathcal{B}$, y que existe una matriz que representa a T en cada base ordenada para V. (En transformaciones de un espacio en otro la matriz depende de dos bases ordenadas, una de V y otra de W.) Para no olvidar esta dependencia se usará la notación

$$[T]_{\mathcal{B}}$$

para la matriz del operador lineal T en la base ordenada $\mathcal{B}$. La manera cómo esta matriz y la base ordenada describen a T, es que para cada α de V

$$[T\alpha]_{\mathcal{B}} = [T]_{\mathcal{B}}[\alpha]_{\mathcal{B}}.$$

Ejemplo 13. Sea V el espacio de las matrices columnas $n \times 1$ sobre el cuerpo F; sea W el espacio de las matrices $m \times 1$ sobre F y sea A una matriz $m \times n$ dada fija sobre F. Sea T la transformación lineal de V en W definida por $T(X) = AX$. Sea $\mathcal{B}$ la base ordenada para V, análoga a la base canónica de F^n; es decir, el i-ésimo vector de $\mathcal{B}$ en la matriz $n \times 1$, X_i tiene un 1 en la fila i y todos los demás elementos 0. Sea $\mathcal{B}'$ la correspondiente base ordenada de W; o sea, el j-ésimo vector de $\mathcal{B}'$ es la matriz $m \times 1$, Y_j, que tiene un 1 en la fila j y todos los demás elementos 0. Entonces la matriz de T respecto al par $\mathcal{B}$, $\mathcal{B}'$ es la misma matriz A. Ello es evidente, pues la matriz AX_j es la j-ésima columna de A.

Ejemplo 14. Sea F un cuerpo y sea T el operador en F^2 definido por

$$T(x_1, x_2) = (x_1, 0).$$

Es fácil ver que T es un operador lineal en F^2. Sea $\mathcal{B}$ la base ordenada canónica de F^2, $\mathcal{B} = \{\epsilon_1, \epsilon_2\}$. Entonces

$$T\epsilon_1 = T(1, 0) = (1, 0) = 1\epsilon_1 + 0\epsilon_2$$
$$T\epsilon_2 = T(0, 1) = (0, 0) = 0\epsilon_1 + 0\epsilon_2$$

con lo que la matriz de T en la base ordenada $\mathcal{B}$ es

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Ejemplo 15. Sea V el espacio de todas las funciones polinomios de R en R de la forma

$$f(x) = c_0 + c_1x + c_2x^2 + c_3x^3$$

esto es, el espacio de las funciones polinomios de grado tres o menor. El operador derivación, del Ejemplo 2, aplica V en V, ya que D «decrece el grado». Sea $\mathcal{B}$ la base ordenada de V que consta de las cuatro funciones f_1, f_2, f_3, f_4 definidas por $f_j(x) = x^{j-1}$. Entonces

$$\begin{aligned} (Df_1)(x) &= 0, & Df_1 &= 0f_1 + 0f_2 + 0f_3 + 0f_4 \\ (Df_2)(x) &= 1, & Df_2 &= 1f_1 + 0f_2 + 0f_3 + 0f_4 \\ (Df_3)(x) &= 2x, & Df_3 &= 0f_1 + 2f_2 + 0f_3 + 0f_4 \\ (Df_4)(x) &= 3x^2, & Df_4 &= 0f_1 + 0f_2 + 3f_3 + 0f_4 \end{aligned}$$

con lo que la matriz de D en la base ordenada $\mathcal{B}$ es

$$[D]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Hemos visto lo que les sucede a las matrices representantes cuando las transformaciones se suman, y es que las matrices se suman. Cabe ahora preguntar qué sucede cuando se componen transformaciones. En forma más precisa, sean V, W y Z espacios vectoriales sobre el cuerpo F de dimensiones n, m y p, respectivamente. Sea T una transformación lineal de V en W y U una transformación lineal de W en Z. Supóngase que se tienen las bases ordenadas

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}, \qquad \mathcal{B}' = \{\beta_1, \ldots, \beta_m\}, \qquad \mathcal{B}'' = \{\gamma_1, \ldots, \gamma_p\}$$

para los respectivos espacios V, W y Z. Sea A la matriz de T respecto al par $\mathcal{B}$, $\mathcal{B}'$ y sea B la matriz de U respecto al par $\mathcal{B}'$, $\mathcal{B}''$. Es fácil ver ahora que la matriz C de la transformación UT respecto al par $\mathcal{B}$, $\mathcal{B}''$ es el producto de B y A; en efecto, si α es cualquier vector de V

$$\begin{aligned} [T\alpha]_{\mathcal{B}'} &= A[\alpha]_{\mathcal{B}} \\ [U(T\alpha)]_{\mathcal{B}''} &= B[T\alpha]_{\mathcal{B}'} \end{aligned}$$

y así

$$[(UT)(\alpha)]_{\mathcal{B}''} = BA[\alpha]_{\mathcal{B}}$$

y luego, por la definición y unicidad de la matriz representante, se debe tener que $C = BA$. Se puede también ver esto haciendo el cálculo

$$\begin{aligned} (UT)(\alpha_j) &= U(T\alpha_j) \\ &= U\left(\sum_{k=1}^{m} A_{kj}\beta_k\right) \\ &= \sum_{k=1}^{m} A_{kj}(U\beta_k) \\ &= \sum_{k=1}^{m} A_{kj} \sum_{i=1}^{p} B_{ik}\gamma_i \\ &= \sum_{i=1}^{p}\left(\sum_{k=1}^{m} B_{ik}A_{kj}\right)\gamma_i \end{aligned}$$

con lo que se debe tener

(3-6) $$C_{ij} = \sum_{k=1}^{m} B_{ik}A_{kj}.$$

Ya se había motivado la definición (3-6) para la multiplicación de matrices por operaciones sobre las filas de una matriz. Aquí se ha visto que una motivación muy convincente para la definición se tiene mediante la composición de transformaciones lineales. Resumiendo, tenemos:

Teorema 13. *Sean V, W y Z espacios vectoriales de dimensión finita sobre el cuerpo F; sea T una transformación lineal de V en W y U una transformación lineal de W en Z. Si $\mathcal{B}$, $\mathcal{B}'$ y $\mathcal{B}''$ son las bases ordenadas de los espacios V, W y Z, respectivamente, y si A es la matriz de T respecto al par $\mathcal{B}$, $\mathcal{B}'$ y B es la matriz de U respecto al par $\mathcal{B}'$, $\mathcal{B}''$, entonces la matriz de la composición UT respecto al par $\mathcal{B}$, $\mathcal{B}''$ es la matriz producto $C = BA$.*

Obsérvese que el Teorema 13 da una demostración de que la multiplicación de matrices es asociativa, demostración que no requiere cálculos y que es independiente de la dada en el Capítulo 1. Debe señalarse también que en el Ejemplo 12 se ha demostrado un caso particular del Teorema 13.

Es importante hacer notar que si T y U son operadores lineales sobre un espacio V y que si se representa por una sola base ordenada $\mathcal{B}$, entonces el Teorema 13 toma la forma simple $[UT]_{\mathcal{B}} = [U]_{\mathcal{B}}[T]_{\mathcal{B}}$. Así, en este caso, la correspondencia que $\mathcal{B}$ determina entre los operadores y las matrices no es solo un isomorfismo del espacio vectorial, sino que también preserva los productos. Una consecuencia inmediata de lo cual es que el operador lineal T es inversible si, y solo si, $[T]_{\mathcal{B}}$ es una matriz inversible. En efecto, el operador identidad I está representado por la matriz identidad en cualquier base ordenada, y así

$$UT = TU = I$$

es equivalente a

$$[U]_{\mathcal{B}}[T]_{\mathcal{B}} = [T]_{\mathcal{B}}[U]_{\mathcal{B}} = I.$$

Además, cuando T es inversible

$$[T^{-1}]_{\mathcal{B}} = [T]_{\mathcal{B}}^{-1}.$$

Quisiéramos ahora investigar qué sucede a la matriz representante cuando se cambia la base ordenada. Por razón de simplicidad se considerará solo el caso de operadores lineales sobre un espacio V, de modo que solo se tenga una base ordenada. Se trata de lo siguiente. Sea T un operador lineal sobre el espacio de dimensión finita V y sean

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{y} \quad \mathcal{B}' = \{\alpha_1', \ldots, \alpha_n'\}$$

dos bases ordenadas para V. ¿De qué manera están relacionadas las matrices $[T]_{\mathcal{B}}$ y $[T]_{\mathcal{B}'}$? Como se observó en el Capítulo 2, existe una matriz (inversible) $n \times n$, P, única tal que

(3-7) $$[\alpha]_{\mathcal{B}} = P[\alpha]_{\mathcal{B}'}$$

para cada vector α en V. Es la matriz $P = [P_1, \ldots, P_n]$ con $P_j = [\alpha_j']_{\mathcal{B}}$. Por definición

(3-8) $$[T\alpha]_{\mathcal{B}} = [T]_{\mathcal{B}}[\alpha]_{\mathcal{B}}.$$

Aplicando (3-7) al vector $T\alpha$, tenemos

(3-9) $$[T\alpha]_{\mathcal{B}} = P[T\alpha]_{\mathcal{B}'}.$$

Combinando (3-7), (3-8) y (3-9) se obtiene

$$[T]_{\mathcal{B}}P[\alpha]_{\mathcal{B}'} = P[T\alpha]_{\mathcal{B}'}$$

o

$$P^{-1}[T]_{\mathcal{B}}P[\alpha]_{\mathcal{B}'} = [T\alpha]_{\mathcal{B}'}$$

con lo que se debe tener que

(3-10) $$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P.$$

Esta es la respuesta del problema.

Antes de enunciar este resultado queremos observar lo siguiente. Existe un único operador lineal U que aplica $\mathcal{B}$ sobre $\mathcal{B}'$, definido por

$$U\alpha_j = \alpha_j', \qquad j = 1, \ldots, n.$$

Este operador U es inversible, ya que aplica una base de V sobre una base de V. La matriz P (anteriormente citada) es precisamente la del operador U en la base ordenada $\mathcal{B}$. En efecto, P está definida por

$$\alpha_j' = \sum_{i=1}^{n} P_{ij}\alpha_i$$

y como $U\alpha_j = \alpha_j'$, esta ecuación puede escribirse

$$U\alpha_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Así $P = [U]_{\mathcal{B}}$ por definición.

Teorema 14. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F, y sean*

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \quad y \quad \mathcal{B}' = \{\alpha_1', \ldots, \alpha_n'\}$$

dos bases ordenadas de V. Supóngase que T es un operador lineal sobre V. Si $P = [P_1, \ldots, P_n]$ es la matriz $n \times n$ de columnas $P_j = [\alpha_j']_{\mathcal{B}}$, entonces

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P.$$

De otra manera, si U es el operador lineal sobre V definido por $U\alpha_j = \alpha_j'$, $j = 1, \ldots, n$, entonces

$$[T]_{\mathcal{B}'} = [U]_{\mathcal{B}}^{-1}[T]_{\mathcal{B}}[U]_{\mathcal{B}}.$$

Ejemplo 16. Sea T el operador lineal sobre R^2 definido por $T(x_1, x_2) = (x_1, 0)$. En el Ejemplo 14 se vio que la matriz de T en la base ordenada canónica $\mathcal{B} = \{\epsilon_1, \epsilon_2\}$ es

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Supóngase que $\mathcal{B}'$ es la base ordenada de R^2, que consta de los vectores $\epsilon_1' = (1, 1)$, $\epsilon_2' = (2, 1)$. Entonces

$$\begin{aligned} \epsilon_1' &= \epsilon_1 + \epsilon_2 \\ \epsilon_2' &= 2\epsilon_1 + \epsilon_2 \end{aligned}$$

de modo que P es la matriz

$$P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}.$$

Después de un breve cálculo se obtiene

$$P^{-1} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}.$$

Con lo que

$$\begin{aligned} [T]_{\mathcal{B}'} &= P^{-1}[T]_{\mathcal{B}}P \\ &= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -1 & -2 \\ 1 & 2 \end{bmatrix}. \end{aligned}$$

Fácilmente se puede comprobar que esto es correcto, pues

$$\begin{aligned} T\epsilon_1' &= (1, 0) = -\epsilon_1' + \epsilon_2' \\ T\epsilon_2' &= (2, 0) = -2\epsilon_1' + 2\epsilon_2'. \end{aligned}$$

Ejemplo 17. Sea V el espacio de las funciones polinomios de R en R de «grado» menor o igual que 3. Como en el Ejemplo 17, sea D el operador derivación en V y sea

$$\mathcal{B} = \{f_1, f_2, f_3, f_4\}$$

la base ordenada de V definida por $f_i(x) = x^{i-1}$. Sea t un número real y defínase $g_i(x) = (x + t)^{i-1}$, esto es,

$$\begin{aligned} g_1 &= f_1 \\ g_2 &= tf_1 + f_2 \\ g_3 &= t^2f_1 + 2tf_2 + f_3 \\ g_4 &= t^3f_1 + 3t^2f_2 + 3tf_3 + f_4. \end{aligned}$$

Como la matriz

$$P = \begin{bmatrix} 1 & t & t^2 & t^3 \\ 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 1 & 3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Se ve fácilmente que es inversible con

$$P^{-1} = \begin{bmatrix} 1 & -t & t^2 & -t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Se sigue que $\mathcal{B}' = \{g_1, g_2, g_3, g_4\}$ es una base ordenada de V. En el Ejemplo 15 se encontró que la matriz de D en la base ordenada es

$$[D]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

La matriz de D en la base ordenada $\mathcal{B}'$ es, pues,

$$\begin{aligned} P^{-1}[D]_{\mathcal{B}}P &= \begin{bmatrix} 1 & -t & t^2 & t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & t & t^2 & t^3 \\ 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 1 & 3t \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -t & t^2 & t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 2 & 6t \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Así D está representada por la misma matriz en las bases ordenadas $\mathcal{B}$ y $\mathcal{B}'$. Por supuesto, que ello puede verse en forma directa, ya que

$$\begin{aligned} Dg_1 &= 0 \\ Dg_2 &= g_1 \\ Dg_3 &= 2g_2 \\ Dg_4 &= 3g_3. \end{aligned}$$

Este ejemplo ilustra algo importante. Si se conoce la matriz de un operador lineal en cierta base $\mathcal{B}$ y se desea encontrar la matriz en otra base ordenada $\mathcal{B}'$, es a menudo más conveniente efectuar el cambio de coordenadas usando la matriz inversible P; pero puede ser mucho más simple encontrar la matriz representante por una aplicación directa de su definición.

Definición. *Sean A y B dos matrices (cuadradas) $n \times n$ sobre el cuerpo F. Se dice que B es* **semejante a** *A* **sobre** *F si existe una matriz inversible $n \times n$, P, sobre F tal que $B = P^{-1}AP$.*

De acuerdo con el Teorema 14, se tiene lo siguiente: Si V es un espacio vectorial de dimensión n sobre el cuerpo F y $\mathcal{B}$ y $\mathcal{B}'$ son dos bases ordenadas de V,

entonces para todo operador lineal T sobre V la matriz $B = [T]_{\mathcal{B}'}$ es semejante a la matriz $A = [T]_{\mathcal{B}}$. El razonamiento también es válido a la inversa. Supóngase que A y B son matrices $n \times n$ y que B es semejante a A. Sea V cualquier espacio de dimensión n sobre F y sea $\mathcal{B}$ una base ordenada de V. Sea T el operador lineal sobre V que está representado en la base $\mathcal{B}$ por A. Si $B = P^{-1}AP$, sea $\mathcal{B}'$ la base ordenada de V obtenida de $\mathcal{B}$ por P, es decir

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Entonces la matriz de T en la base ordenada $\mathcal{B}'$ será B.

Así la afirmación de que B es semejante a A quiere decir que en cada espacio de dimensión n sobre F las matrices A y B representan la misma transformación lineal en las dos (posibles) bases ordenadas diferentes.

Obsérvese que toda matriz $n \times n$, A, es semejante a sí misma, tomando $P = I$; si B es semejante a A, entonces A es semejante a B, ya que $B = P^{-1}AP$ implica que $A = (P^{-1})^{-1}BP^{-1}$; si B es semejante a A y C es semejante a B, entonces C es semejante a A, pues $B = P^{-1}AP$ y $C = Q^{-1}BQ$ implica que $C = (PQ)^{-1}A(PQ)$. Así, pues, la semejanza es una relación de equivalencia en el conjunto de las matrices $n \times n$ sobre el cuerpo F. Obsérvese también que la única matriz semejante a la matriz identidad I es I y que la única matriz semejante a la matriz nula es la misma matriz nula.

Ejercicios

1. Sea T el operador lineal sobre C^2 definido por $T(x_1, x_2) = (x_1, 0)$. Sea $\mathcal{B}$ la base ordenada canónica de C^2 y sea $\mathcal{B}' = \{\alpha_1, \alpha_2\}$ la base ordenada definida por $\alpha_1 = (1, i)$, $\alpha_2 = (-i, 2)$.

(a) ¿Cuál es la matriz de T respecto al par $\mathcal{B}$, $\mathcal{B}'$?
(b) ¿Cuál es la matriz de T respecto al par $\mathcal{B}'$, $\mathcal{B}$?
(c) ¿Cuál es la matriz de T en la base ordenada $\mathcal{B}'$?
(d) ¿Cuál es la matriz de T en la base ordenada $\{\alpha_2, \alpha_1\}$?

2. Sea T la transformación lineal de R^3 en R^2 definida por

$$T(x_1, x_2, x_3) = (x_1 + x_2, 2x_3 - x_1).$$

(a) Si $\mathcal{B}$ es la base ordenada canónica de R^3 y $\mathcal{B}'$ es la base ordenada canónica de R^2, ¿cuál es la matriz de T respecto al par $\mathcal{B}$, $\mathcal{B}'$?

(b) Si $\mathcal{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ y $\mathcal{B}' = \{\beta_1, \beta_2\}$, donde

$$\alpha_1 = (1, 0, -1), \quad \alpha_2 = (1, 1, 1), \quad \alpha_3 = (1, 0, 0), \quad \beta_1 = (0, 1), \quad \beta_2 = (1, 0)$$

¿Cuál es la matriz de T respecto al par $\mathcal{B}$, $\mathcal{B}'$?

3. Sea T un operador lineal sobre F^n, sea A la matriz de T en la base ordenada canónica para F^n y sea W el subespacio de F^n generado por los vectores columna de A. ¿Qué relación existe entre W y T?

4. Sea V un espacio vectorial bidimensional sobre el cuerpo F y sea $\mathcal{B}$ una base ordenada de V. Si T es un operador lineal en V y

$$[T]_{\mathcal{B}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

demostrar que $T^2 - (a + d)T + (ad - bc)I = 0$.

5. Sea T el operador lineal sobre R^3 cuya matriz en la base canónica es

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 3 & 4 \end{bmatrix}.$$

Encontrar una base de la imagen de T y una base del espacio nulo de T.

6. Sea T el operador lineal sobre R^2 definido por

$$T(x_1, x_2) = (-x_2, x_1).$$

(a) ¿Cuál es la matriz de T en la base ordenada canónica de R^2?
(b) ¿Cuál es la matriz de T en la base ordenada $\mathcal{B} = \{\alpha_1, \alpha_2\}$, donde $\alpha_1 = (1, 2)$ y $\alpha_2 = (1, -1)$?
(c) Demostrar que para cada número real c el operador $(T - cI)$ es inversible.
(d) Demostrar que si $\mathcal{B}$ es cualquier base ordenada para R^2 y $[T]_{\mathcal{B}} = A$, entonces $A_{12}A_{21} \neq 0$.

7. Sea T el operador lineal en R^3 definido por

$$T(x_1, x_2, x_3) = (3x_1 + x_3, -2x_1 + x_2, -x_1 + 2x_2 + 4x_3).$$

(a) ¿Cuál es la matriz de T en la base ordenada canónica de R^3?
(b) ¿Cuál es la matriz de T en la base ordenada

$$\{\alpha_1, \alpha_2, \alpha_3\}$$

donde $\alpha_1 = (1, 0, 1)$, $\alpha_2 = (-1, 2, 1)$ y $\alpha_3 = (2, 1, 1)$?
(c) Demostrar que T es inversible y dar una expresión de T^{-1} tal como la que define a T.

8. Sea θ un número real. Demostrar que las dos matrices siguientes son similares sobre el cuerpo de los números complejos:

$$\begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}, \qquad \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$$

(*Sugerencia*: Sea T el operador lineal sobre C^2 representado por la primera matriz en la base ordenada canónica. Encontrar entonces vectores α_1 y α_2 tales que $T\alpha_1 = e^{i\theta}\alpha_1$, $T\alpha_2 = e^{-i\theta}\alpha_2$ y $\{\alpha_1, \alpha_2\}$ sea una base.)

9. Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sean S y T operadores lineales sobre V. Se pregunta: ¿cuándo existen dos bases ordenadas $\mathcal{B}$ y $\mathcal{B}'$ de V tales que $[S]_{\mathcal{B}} = [T]_{\mathcal{B}'}$? Demostrar que tales bases existen si, y solo si, existe un operador lineal inversible U en V tal que $T = USU^{-1}$. (*Esquema de la demostración*: Si $[S]_{\mathcal{B}} = [T]_{\mathcal{B}'}$, sea U el operador que aplica $\mathcal{B}$ sobre $\mathcal{B}'$ y demostrar que $S = UTU^{-1}$. Recíprocamente, si $T = USU^{-1}$ para algún U inversible, sea $\mathcal{B}$ cualquier base ordenada para V y sea $\mathcal{B}'$ su imagen por U. Demostrar entonces que $[S]_{\mathcal{B}} = [T]_{\mathcal{B}'}$.)

10. Se sabe que el operador lineal T sobre R^2 definido por $T(x_1, x_2) = (x_1, 0)$ está representado, en la base ordenada canónica, por la matriz

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Este operador cumple $T^2 = T$. Demostrar que si S es un operador lineal sobre R^2 tal que $S^2 = S$, entonces $S = 0$, o $S = I$, o existe una base ordenada $\mathcal{B}$ de R^2 tal que $[S]_{\mathcal{B}} = A$ (anteriormente).

11. Sea W el espacio de todas las matrices columna $n \times 1$ sobre el cuerpo F. Si A es una matriz $n \times n$ sobre F, entonces A define un operador lineal L_A sobre W por la multiplicación a la izquierda $L_A(X) = AX$. Demostrar que cada operador lineal sobre W es el producto a la izquierda de alguna matriz $n \times n$; es decir, es L_A para algún A.

Ahora supóngase que V es un espacio vectorial de dimensión n sobre el cuerpo F y que $\mathcal{B}$ es una base ordenada de V. Para cada α de V se define $U\alpha = [\alpha]_{\mathcal{B}}$. Demostrar que U es un isomorfismo de V sobre W. Si T es un operador lineal sobre V, entonces UTU^{-1} es un operador lineal sobre W. Por tanto, UTU^{-1} es el producto a la izquierda por alguna matriz $n \times n$, A. ¿Cuál es A?

12. Sea V un espacio vectorial de dimensión n sobre el cuerpo F y sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ordenada de V.

(a) Conforme al Teorema 1 existe un operador lineal único T sobre V tal que

$$T\alpha_j = \alpha_{j+1}, \qquad j = 1, \ldots, n-1, \qquad T\alpha_n = 0.$$

¿Cuál es la matriz A de T en la base ordenada $\mathcal{B}$?

(b) Demostrar que $T^n = 0$, pero que $T^{n-1} \neq 0$.

(c) Sea S cualquier operador lineal en V tal que $S^n = 0$, pero con $S^{n-1} \neq 0$. Demostrar que existe una base ordenada $\mathcal{B}'$ de V tal que la matriz de S en la base ordenada $\mathcal{B}'$ es la matriz A de la parte (a).

(d) Demostrar que si M y N son matrices $n \times n$ sobre F tales que $M^n = N^n = 0$, con $M^{n-1} \neq 0 \neq N^{n-1}$, entonces M y N son semejantes.

13. Sean V y W espacios vectoriales de dimensión finita sobre el cuerpo F y sea T una transformación lineal de V en W. Si

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{y} \quad \mathcal{B}' = \{\beta_1, \ldots, \beta_m\}$$

son bases ordenadas de V y W, respectivamente, defínase la transformación lineal $E^{p,q}$ como en la demostración del Teorema 5: $E^{p,q}(\alpha_i) = \delta_{iq}\beta_p$. Entonces, los $E^{p,q}$, $1 \leq p \leq m$, $1 \leq q \leq n$, forman una base de $L(V, W)$, y así

$$T = \sum_{p=1}^{m} \sum_{q=1}^{n} A_{pq}E^{p,q}$$

para ciertos escalares $A_{p,q}$ (las coordenadas de T en esta base de $L(V, W)$). Demostrar que la matriz A con elementos $A(p, q) = A_{pq}$ es justamente la matriz de T respecto al par $\mathcal{B}$, $\mathcal{B}'$.

3.5. Funcionales lineales

Si V es un espacio vectorial sobre el cuerpo F, una transformación lineal f de V en el cuerpo de escalares F se llama también un **funcional lineal** sobre V. Si se comienza desde el principio, esto quiere decir que f es una función de V en F, tal que

$$f(c\alpha + \beta) = cf(\alpha) + f(\beta)$$

para todos los vectores α y β de V y todos los escalares c de F. El concepto de funcional lineal es importante para el estudio de los espacios vectoriales de dimensión finita, pues ayuda a organizar y clarificar el estudio de los subespacios, las ecuaciones lineales y las coordenadas.

Ejemplo 18. Sea F un cuerpo y sean $a_1, \ldots, a_n$ escalares pertenecientes a F. Defínase una función f en F^n por

$$f(x_1, \ldots, x_n) = a_1x_1 + \cdots + a_nx_n.$$

Entonces f es un funcional lineal sobre F^n. Es el funcional lineal representado por la matriz $[a_1, \ldots, a_n]$ respecto a la base ordenada canónica de F^n y la base $\{1\}$ de F:

$$a_j = f(\epsilon_j), \qquad j = 1, \ldots, n.$$

Todo funcional lineal sobre F^n es de esta forma para ciertos escalares $a_1, \ldots, a_n$. Ello se sigue de la definición de funcional lineal, ya que definimos $a_j = f(\epsilon_j)$ y empleando la linealidad

$$\begin{aligned} f(x_1, \ldots, x_n) &= f\left(\sum_j x_j\epsilon_j\right) \\ &= \sum_j x_jf(\epsilon_j) \\ &= \sum_j a_jx_j. \end{aligned}$$

Ejemplo 19. He aquí un ejemplo importante de un funcional lineal. Sea n un entero positivo y F un cuerpo. Si A es una matriz $n \times n$ con elementos en F, la **traza** de A es el escalar

$$\operatorname{tr} A = A_{11} + A_{22} + \cdots + A_{nn}.$$

La función traza es un funcional lineal en el espacio de las matrices $F^{n\times n}$, pues

$$\begin{aligned} \operatorname{tr}(cA + B) &= \sum_{i=1}^{n} (cA_{ii} + B_{ii}) \\ &= c\sum_{i=1}^{n} A_{ii} + \sum_{i=1}^{n} B_{ii} \\ &= c \operatorname{tr} A + \operatorname{tr} B. \end{aligned}$$

Ejemplo 20. Sea V el espacio de todas las funciones polinomios del cuerpo F en sí mismo. Sea t un elemento de F. Si se define

$$L_t(p) = p(t)$$

entonces L_t es un funcional lineal en V. Es usual describir esto diciendo que, para cada t, la «valuación en t» es un funcional lineal en el espacio de las funciones polinomios. Habría que advertir que el que las funciones sean polinomios no interviene en el ejemplo. La valuación en t es un funcional lineal en el espacio de todas las funciones de F en F.

Ejemplo 21. Este puede ser el funcional lineal más importante en matemática. Sea $[a, b]$ un intervalo cerrado del eje real y sea $C([a, b])$ el espacio de las funciones reales continuas sobre $[a, b]$. Entonces

$$L(g) = \int_a^b g(t)\, dt$$

define un funcional lineal L en $C([a, b])$.

Si V es un espacio vectorial, el conjunto de todos los funcionales lineales sobre V forman, naturalmente, un espacio vectorial. Es el espacio $L(V, F)$. Se designa este espacio por V^* y se le llama **espacio dual** del V:

$$V^* = L(V, F).$$

Si V es de dimensión finita se puede obtener una descripción muy explícita del espacio dual V^*. Por el Teorema 5 sabemos algo acerca del espacio V^*:

$$\dim V^* = \dim V.$$

Sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base de V. Conforme al Teorema 1 existe (para cada i) un funcional lineal único f_i en V tal que

(3-11) $$f_i(\alpha_j) = \delta_{ij}.$$

De esta forma se obtiene de $\mathcal{B}$ un conjunto de n funcionales lineales distintos $f_1, \ldots, f_n$ sobre V. Estos funcionales son también linealmente independientes, pues supóngase que

(3-12) $$f = \sum_{i=1}^{n} c_i f_i.$$

Entonces

$$\begin{aligned} f(\alpha_j) &= \sum_{i=1}^{n} c_i f_i(\alpha_j) \\ &= \sum_{i=1}^{n} c_i \delta_{ij} \\ &= c_j. \end{aligned}$$

En particular, si f es el funcional cero, $f(\alpha_j) = 0$ para cada j y, por tanto, los escalares c_j son todos ceros. Entonces los $f_1, \ldots, f_n$ son n funcionales linealmente independientes, y como se sabe que V^* tiene dimensión n, deben ser tales que $\mathcal{B}^* = \{f_1, \ldots, f_n\}$ es una base de V^*. Esta base se llama **base dual** de $\mathcal{B}$.

Teorema 15. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base de V. Entonces existe una única base dual $\mathcal{B}^* = \{f_1, \ldots, f_n\}$ de V^* tal que $f_i(\alpha_j) = \delta_{ij}$. Para cada funcional lineal f sobre V se tiene*

(3-13) $$f = \sum_{i=1}^{n} f(\alpha_i) f_i$$

y para cada vector α de V se tiene

(3-14) $$\alpha = \sum_{i=1}^{n} f_i(\alpha)\alpha_i.$$

Demostración. Se ha visto antes que existe una base única que es «dual» de $\mathcal{B}$. Si f es un funcional lineal sobre V, entonces f es una combinación lineal

(3-12) de los f_i y, como observamos después de (3-12), los escalares c_j vienen dados por $c_j = f(\alpha_j)$. En forma análoga, si

$$\alpha = \sum_{i=1}^{n} x_i\alpha_i$$

es un vector de V, entonces

$$\begin{aligned} f_j(\alpha) &= \sum_{i=1}^{n} x_i f_j(\alpha_i) \\ &= \sum_{i=1}^{n} x_i\delta_{ij} \\ &= x_j \end{aligned}$$

de modo que la expresión única para α, como combinación lineal de los α_i, es

$$\alpha = \sum_{i=1}^{n} f_i(\alpha)\alpha_i. \quad \blacksquare$$

La ecuación (3-14) suministra un buen modo de describir qué es la base dual. Dice que, si $\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}$ es una base ordenada de V y $\mathfrak{B}^* = \{f_1, \ldots, f_n\}$ es la base dual, entonces f_i es precisamente la función que asigna a cada vector α en V la i-ésima coordenada de α respecto a la base ordenada $\mathfrak{B}$. Así que también se pueden llamar los f_i funciones coordenadas de $\mathfrak{B}$. La fórmula (3-13), cuando se combina con (3-14), dice que: Si f pertenece a V^* y si $f(\alpha_i) = a_i$, entonces si

$$\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n$$

se tiene

(3-15) $$f(\alpha) = a_1x_1 + \cdots + a_nx_n.$$

En otras palabras, si se elige una base ordenada $\mathfrak{B}$ de V y se expresa cada vector en V por su n-tuple de coordenadas $(x_1, \ldots, x_n)$ respecto a $\mathfrak{B}$, entonces cada funcional lineal en V tiene la forma (3-15). Esta es la generalización natural del Ejemplo 18, que es el caso especial en que $V = F^n$ y $\mathfrak{B} = \{\epsilon_1, \ldots, \epsilon_n\}$.

Ejemplo 22. Sea V el espacio vectorial de todas las funciones polinomios de R en R que tienen grado menor o igual que 2. Sean t_1, t_2, t_3 tres números reales *distintos* arbitrarios, y sea

$$L_i(p) = p(t_i).$$

Entonces L_1, L_2 y L_3 son funcionales lineales sobre V. Estos funcionales son linealmente independientes; en efecto, supóngase que

$$L = c_1L_1 + c_2L_2 + c_3L_3.$$

Si $L = 0$; es decir, si $L(p) = 0$ para todo p de V, entonces aplicando L a las «funciones» polinomios particulares 1, x, x^2, tenemos

$$\begin{aligned} c_1 + c_2 + c_3 &= 0 \\ t_1c_1 + t_2c_2 + t_3c_3 &= 0 \\ t_1^2c_1 + t_2^2c_2 + t_3^2c_3 &= 0 \end{aligned}$$

De aquí se sigue que $c_1 = c_2 = c_3 = 0$, porque (como lo indica un ligero cálculo) la matriz

$$\begin{bmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \\ t_1^2 & t_2^2 & t_3^2 \end{bmatrix}$$

es inversible cuando t_1, t_2 y t_3 son distintos. Entonces los L_i son independientes, y como V tiene dimensión 3, estos funcionales forman una base de V^*. ¿Cuál es la base de V de la que ésta es la dual? Tal base $\{p_1, p_2, p_3\}$ de V debe satisfacer

$$L_i(p_j) = \delta_{ij}$$

o

$$p_j(t_i) = \delta_{ij}.$$

Estas funciones polinomios son, como se ve bien fácilmente,

$$p_1(x) = \frac{(x - t_2)(x - t_3)}{(t_1 - t_2)(t_1 - t_3)}$$

$$p_2(x) = \frac{(x - t_1)(x - t_3)}{(t_2 - t_1)(t_2 - t_3)}$$

$$p_3(x) = \frac{(x - t_1)(x - t_2)}{(t_3 - t_1)(t_3 - t_2)}.$$

La base $\{p_1, p_2, p_3\}$ para V es interesante, ya que de acuerdo con (3-14) se tiene para cada p en V

$$p = p(t_1)p_1 + p(t_2)p_2 + p(t_3)p_3.$$

Con lo que, si c_1, c_2 y c_3 son números reales cualesquiera, existe exactamente una función polinomio p sobre R que tiene grado a lo más 2 y satisface a $p(t_j) = c_j$, $j = 1, 2, 3$. Esta función polinomio es $p = c_1p_1 + c_2p_2 + c_3p_3$.

Sea estudiar ahora la relación entre los funcionales lineales y los subespacios. Si f es un funcional lineal no nulo, entonces el rango de f es 1, ya que la imagen de f es un subespacio no nulo del cuerpo escalar y debe ser (por tanto) el cuerpo escalar. Si el espacio de referencia V es de dimensión finita, el rango más la nulidad (Teorema 2) nos dice que el espacio nulo N_f tiene dimensión

$$\dim N_f = \dim V - 1.$$

En un espacio vectorial de dimensión n, un subespacio de dimensión $n - 1$ se llama un **hiperespacio.** Tales espacios se llaman también a veces hiperplanos o subespacios de codimensión 1. ¿Es todo hiperespacio el espacio nulo de un funcional lineal? Se ve fácilmente que la respuesta es sí. No es mucho más difícil demostrar que todo subespacio de dimensión d es la intersección de los espacios nulos de $(n - d)$ funciones lineales (Teorema 16, más adelante).

Definición. *Si V es un espacio vectorial sobre el cuerpo F y S es un subconjunto de V, el* **anulador** *de S es el conjunto S^0 de funcionales lineales f sobre V tales que $f(\alpha) = 0$ para todo α de S.*

Debería ser claro para el lector que S^0 es un subespacio de V^*, sea o no S un subespacio de V. Si S es el conjunto que consta del solo vector cero, entonces $S^0 = V^*$. Si $S = V$, entonces S^0 es el subespacio cero de V^*. (Esto es fácil de ver cuando V es de dimensión finita.)

Teorema 16. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea W un subespacio de V. Entonces*

$$\dim W + \dim W^0 = \dim V.$$

Demostración. Sea k la dimensión de W y $\{\alpha_1, \ldots, \alpha_k\}$ una base de W. Se eligen vectores $\alpha_{k+1}, \ldots, \alpha_n$ en V, de modo que $\{\alpha_1, \ldots, \alpha_n\}$ sea una base para V. Sea $\{f_1, \ldots, f_n\}$ la base de V^* que es dual de esta base de V. Se afirma que $\{f_{k+1}, \ldots, f_n\}$ es una base del anulador W^0. Efectivamente, f_i pertenece a W^0 para $i \geq k + 1$, ya que

$$f_i(\alpha_j) = \delta_{ij}$$

y $\delta_{ij} = 0$ si $i \geq k + 1$ y $j \leq k$; de esto se sigue que, para $i \geq k + 1$, $f_i(\alpha) = 0$ siempre que α sea una combinación lineal de $\alpha_1, \ldots, \alpha_n$. Los funcionales $f_{k+1}, \ldots, f_n$ son independientes, de modo que todo lo que se debe demostrar es que generan W^0. Supóngase que f está en V^*. Entonces

$$f = \sum_{i=1}^{n} f(\alpha_i)f_i$$

de modo que si f está en W^0 se tiene $f(\alpha_i) = 0$ para $i \leq k$ y

$$f = \sum_{i=k+1}^{n} f(\alpha_i)f_i.$$

Hemos mostrado que si $\dim W = k$ y $\dim V = n$, entonces $\dim W^0 = n - k$. ▮

Corolario. *Si W es un subespacio de dimensión k de un espacio vectorial V de dimensión n, entonces W es la intersección de $(n - k)$ hiperespacios en V.*

Demostración. Este es un corolario de la demostración del Teorema 16 más bien que su enunciado. Con la nomenclatura de la demostración, W es justamente el conjunto de los vectores α tales que $f_i(\alpha) = 0$, $i = k + 1, \ldots, n$. En caso de que $k = n - 1$, W es el espacio nulo de f_n. ▮

Corolario. *Si W_1 y W_2 son subespacios de un espacio vectorial de dimensión finita, entonces $W_1 = W_2$ si, y solo si, $W_1^0 = W_2^0$.*

Demostración. Si $W_1 = W_2$, entonces es claro que $W_1^0 = W_2^0$. Si $W_1 \neq W_2$, entonces uno de los dos subespacios contiene un vector que no está en el otro. Supóngase que hay un vector α que está en W_2, pero no en W_1. Por el corolario anterior (o por la demostración del Teorema 16), existe un funcional lineal f tal que $f(\beta) = 0$ para todo β de W_1, pero $f(\alpha) \neq 0$. Entonces f está en W_1^0, pero no en W_2^0 y $W_1^0 \neq W_2^0$. ▮

En la sección siguiente se darán demostraciones distintas para estos dos corolarios. El primer corolario dice que, si se elige una base ordenada del espacio, cada subespacio de dimensión k puede describirse dando $(n - k)$ condiciones lineales homogéneas para las coordenadas respecto de esa base.

Obsérvese, ahora, el sistema de ecuaciones lineales homogéneas desde el punto de vista de los funcionales lineales. Supóngase que se tiene un sistema lineal de ecuaciones

$$\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= 0 \\ \vdots \qquad\quad & \vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &= 0 \end{aligned}$$

del que se quiere encontrar sus soluciones. Si se designa por f_i, $i = 1, \ldots, m$ al funcional lineal en F^n definido por

$$f_i(x_1, \ldots, x_n) = A_{i1}x_1 + \cdots + A_{in}x_n$$

entonces se busca el subespacio de F^n para todos los α tales que

$$f_i(\alpha) = 0, \qquad i = 1, \ldots, m.$$

En otras palabras, se busca el subespacio anulado por $f_1, \ldots, f_m$. La reducción por filas de la matriz de los coeficientes da un método sistemático para encontrar este subespacio. El n-tuple $(A_{i1}, \ldots, A_{in})$ da las coordenadas del funcional lineal f_i respecto a la base que es dual a la base canónica de F^n. El espacio de las filas de la matriz de los coeficientes puede, pues, considerarse como el espacio de los funcionales lineales generado por $f_1, \ldots, f_m$. El espacio solución es el subespacio anulado por este espacio de funcionales.

Ahora se puede considerar el sistema de ecuaciones desde el punto de vista «dual». Esto es, supóngase que se den m vectores de F^n

$$\alpha_i = (A_{i1}, \ldots, A_{in})$$

y que se desea hallar el anulador del subespacio generado por estos vectores. Como un funcional lineal típico sobre F^n tiene la forma

$$f(x_1, \ldots, x_n) = c_1x_1 + \cdots + c_nx_n$$

la condición de que f esté en este anulador es que

$$\sum_{j=1}^{n} A_{ij}c_j = 0, \qquad i = 1, \ldots, m$$

esto es, que $(c_1, \ldots, c_n)$ sea una solución del sistema $AX = 0$. Desde este punto de vista, la reducción por filas da un método sistemático para hallar el anulador del subespacio generado por un conjunto finito de vectores de F^n.

Ejemplo 23. He aquí tres funcionales lineales sobre R^4:

$$\begin{aligned} f_1(x_1, x_2, x_3, x_4) &= x_1 + 2x_2 + 2x_3 + x_4 \\ f_2(x_1, x_2, x_3, x_4) &= 2x_2 + x_4 \\ f_3(x_1, x_2, x_3, x_4) &= -2x_1 - 4x_3 + 3x_4. \end{aligned}$$

El subespacio que anulan puede ser encontrado explícitamente hallando la matriz escalón reducida por filas de la matriz

$$A = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ -2 & 0 & -4 & 3 \end{bmatrix}.$$

Un cálculo breve, o una observación del Ejemplo 21 del Capítulo 2, muestra que

$$R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Por tanto, los funcionales lineales

$$\begin{aligned} g_1(x_1, x_2, x_3, x_4) &= x_1 + 2x_3 \\ g_2(x_1, x_2, x_3, x_4) &= x_2 \\ g_3(x_1, x_2, x_3, x_4) &= x_4 \end{aligned}$$

generan el mismo subespacio de $(R^4)^*$ y anulan el mismo subespacio de R^4 como lo hacen f_1, f_2, f_3. El subespacio anulado consta de los vectores con

$$\begin{aligned} x_1 &= -2x_3 \\ x_2 &= x_4 = 0. \end{aligned}$$

Ejemplo 24. Sea W el subespacio de R^5 generado por los vectores

$$\begin{aligned} \alpha_1 &= (2, -2, 3, 4, -1), & \alpha_3 &= (0, 0, -1, -2, 3) \\ \alpha_2 &= (-1, 1, 2, 5, 2), & \alpha_4 &= (1, -1, 2, 3, 0). \end{aligned}$$

¿Cómo se puede describir W^0, anulador de W? Formando la matriz 4×5, A, con vectores filas α_1, α_2, α_3, α_4, y se encuentra la matriz escalón reducida por filas R que es equivalente por filas a A:

$$A = \begin{bmatrix} 2 & -2 & 3 & 4 & -1 \\ -1 & 1 & 2 & 5 & 2 \\ 0 & 0 & -1 & -2 & 3 \\ 1 & -1 & 2 & 3 & 0 \end{bmatrix} \longrightarrow R = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Si f es un funcional lineal sobre R^5:

$$f(x_1, \ldots, x_5) = \sum_{j=1}^{5} c_j x_j$$

entonces f está en W^0 si, y solo si, $f(\alpha_i) = 0$, $i = 1, 2, 3, 4$; es decir, si, y solo si,

$$\sum_{j=1}^{5} A_{ij} c_j = 0, \qquad 1 \leq i \leq 4.$$

Esto es equivalente a

$$\sum_{j=1}^{5} R_{ij} c_j = 0, \qquad 1 \leq i \leq 3$$

o

$$\begin{aligned} c_1 - c_2 - c_4 &= 0 \\ c_3 + 2c_4 &= 0 \\ c_5 &= 0. \end{aligned}$$

Todos estos funcionales lineales f se obtienen asignando valores arbitrarios a c_2 y c_4, por ejemplo, $c_2 = a$ y $c_4 = b$, y a continuación hallando los correspondientes $c_1 = a + b$, $c_3 = -2b$, $c_5 = 0$.
Con lo que W^0 consta de los funcionales lineales f de la forma

$$f(x_1, x_2, x_3, x_4, x_5) = (a + b)x_1 + ax_2 - 2bx_3 + bx_4.$$

La dimensión de W^0 es 2, y una base $\{f_1, f_2\}$ para W^0 puede encontrarse primeramente tomando $a = 1$, $b = 0$, y entonces $a = 0$, $b = 1$:

$$f_1(x_1, \ldots, x_5) = x_1 + x_2$$
$$f_2(x_1, \ldots, x_5) = x_1 - 2x_3 + x_4.$$

El f anterior general en W^0 es $f = af_1 + bf_2$.

Ejercicios

1. En R^3, sea $\alpha_1 = (1, 0, 1)$, $\alpha_2 = (0, 1, -2)$, $\alpha_3 = (-1, -1, 0)$
(a) Si f es un funcional lineal sobre R^3 tal que

$$f(\alpha_1) = 1, \qquad f(\alpha_2) = -1, \qquad f(\alpha_3) = 3,$$

y si $\alpha = (a, b, c)$, hallar $f(\alpha)$.
(b) Describir explícitamente un funcional lineal f sobre R^3 tal que

$$f(\alpha_1) = f(\alpha_2) = 0 \text{ pero } f(\alpha_3) \neq 0.$$

(c) Sea f cualquier funcional lineal tal que

$$f(\alpha_1) = f(\alpha_2) = 0 \quad \text{y} \quad f(\alpha_3) \neq 0.$$

Si $\alpha = (2, 3, -1)$, demostrar que $f(\alpha) \neq 0$.

2. Sea $\mathcal{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ la base para C^3 definida por

$$\alpha_1 = (1, 0, -1), \qquad \alpha_2 = (1, 1, 1), \qquad \alpha_3 = (2, 2, 0).$$

Hallar la base dual de $\mathcal{B}$.

3. Si A y B son matrices $n \times n$ sobre el cuerpo F, demostrar que traza (AB) = traza (BA). Entonces demostrar que las matrices semejantes tienen la misma traza.

4. Sea V el espacio vectorial de todas las funciones polinomios p de R en R que tienen grado 2 o menor:

$$p(x) = c_0 + c_1x + c_2x^2.$$

Se definen tres funciones lineales sobre V por

$$f_1(p) = \int_0^1 p(x)\,dx, \qquad f_2(p) = \int_0^2 p(x)\,dx, \qquad f_3(p) = \int_0^{-1} p(x)\,dx.$$

Demostrar que $\{f_1, f_2, f_3\}$ es una base de V^*, presentando la base de V de la cual ésta es dual.

5. Si A y B son matrices complejas $n \times n$, hacer ver que es imposible que $AB - BA = I$.

6. Sean m y n enteros positivos y F un cuerpo. Sean $f_1, \ldots, f_m$ funcionales lineales en F^n. Para α en F^n se define

$$T\alpha = (f_1(\alpha), \ldots, f_m(\alpha)).$$

Demostrar que T es una transformación lineal de F^n en F^m. Demostrar luego que toda transformación lineal de F^n en F^m es de la forma anterior para ciertos $f_1, \ldots, f_m$.

7. Sea $\alpha_1 = (1, 0, -1, 2)$ y $\alpha_2 = (2, 3, 1, 1)$ y sea W el subespacio de R^4 generado por α_1 y α_2. ¿Qué funcionales lineales f:

$$f(x_1, x_2, x_3, x_4) = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4$$

están en el anulador de W?

8. Sea W el subespacio de R^5 generado por los vectores

$$\alpha_1 = \epsilon_1 + 2\epsilon_2 + \epsilon_3, \qquad \alpha_2 = \epsilon_2 + 3\epsilon_3 + 3\epsilon_4 + \epsilon_5$$
$$\alpha_3 = \epsilon_1 + 4\epsilon_2 + 6\epsilon_3 + 4\epsilon_4 + \epsilon_5.$$

Hallar una base de W^0.

9. Sea V el espacio vectorial de todas las matrices 2×2 sobre el cuerpo de los números reales, y sea

$$B = \begin{bmatrix} 2 & -2 \\ -1 & 1 \end{bmatrix}.$$

Sea W el subespacio de V que consta de todas las A tales que $AB = 0$. Sea f un funcional lineal sobre V que está en el anulador de W. Supóngase que $f(I) = 0$ y $f(C) = 3$, donde I es la matriz identidad 2×2 y

$$C = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hallar $f(B)$.

10. Sea F un subcuerpo de los números complejos. Se definen n funcionales lineales en F^n $(n \geq 2)$ por

$$f_k(x_1, \ldots, x_n) = \sum_{j=1}^{n} (k - j)x_j, \qquad 1 \leq k \leq n.$$

¿Cuál es la dimensión del subespacio anulado por $f_1, \ldots, f_n$?

11. Sean W_1 y W_2 subespacios de un espacio vectorial V de dimensión finita.
   (a) Demostrar que $(W_1 + W_2)^0 = W_1^0 \cap W_2^0$.
   (b) Demostrar que $(W_1 \cap W_2)^0 = W_1^0 + W_2^0$.

12. Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea W un subespacio de V. Si f es un funcional lineal sobre W, demostrar que existe un funcional lineal g sobre V tal que $g(\alpha) = f(\alpha)$ para todo α del subespacio W.

13. Sea F un subcuerpo del cuerpo de los números complejos y sea V cualquier espacio vectorial sobre F. Supóngase que f y g son funcionales lineales sobre V tales que la función h definida por $h(\alpha) = f(\alpha)g(\alpha)$ sea también un funcional lineal sobre V. Demostrar que $f = 0$ o $g = 0$.

14. Sea F un cuerpo de característica cero y sea V un espacio vectorial de dimensión finita sobre F. Si $\alpha_1, \ldots, \alpha_m$ son un número finito de vectores de V, distintos del vector nulo, demostrar que existe un funcional lineal f sobre V tal que

$$f(\alpha_i) \neq 0, \qquad i = 1, \ldots, m.$$

15. De acuerdo con el Ejercicio 3, las matrices semejantes tienen la misma traza. Entonces se puede definir la traza de un operador lineal en un espacio de dimensión finita como la traza de cualquier matriz que represente al operador en una base ordenada. Esta está bien definida, ya que todas las matrices que representan un operador son semejantes.

Sea ahora V el espacio de todas las matrices 2×2 sobre el cuerpo F y sea P una matriz 2×2 dada. Sea T el operador lineal sobre V definido por $T(A) = PA$. Demostrar que traza $(T) = 2$ traza (P).

16. Demostrar que el funcional traza de una matriz $n \times n$ es único en el siguiente sentido. Si W es el espacio de las matrices $n \times n$ sobre el cuerpo F y si f es un funcional lineal sobre W tal que $f(AB) = f(BA)$ para todo A y B de W, entonces f es un múltiplo escalar de la función traza. Si además $f(I) = n$, entonces f es la función traza.

17. Sea W el espacio de las matrices $n \times n$ sobre el cuerpo F y sea W_0 el subespacio generado por las matrices C de la forma $C = AB - BA$. Demostrar que W_0 es exactamente el subespacio de las matrices que tienen traza cero. (*Indicación*: ¿Cuál es la dimensión del espacio de las matrices de traza cero? Considerar las matrices «básicas», es decir, las matrices que tienen solo un elemento no nulo, para construir suficientes matrices linealmente independientes de la forma $AB - BA$.)

3.6. El doble dual

Una pregunta respecto a bases duales que no se contestó en la sección anterior, era si toda base de V^* es la dual de alguna base de V. Una posibilidad de contestar esta pregunta es considerar V^{**}, espacio dual de V^*.

Si α es un vector de V, entonces α induce un funcional lineal L_α sobre V^*, definido por

$$L_\alpha(f) = f(\alpha), \qquad f \text{ en } V^*.$$

El hecho de que L_α sea lineal no es más que una reformulación de la definición de las operaciones lineales en V^*:

$$\begin{aligned} L_\alpha(cf + g) &= (cf + g)(\alpha) \\ &= (cf)(\alpha) + g(\alpha) \\ &= cf(\alpha) + g(\alpha) \\ &= cL_\alpha(f) + L_\alpha(g). \end{aligned}$$

Si V es de dimensión finita y $\alpha \neq 0$, entonces $L_\alpha \neq 0$; en otras palabras, existe un funcional lineal f tal que $f(\alpha) \neq 0$. La demostración es muy simple y fue dada en la Sección 3.5: Elíjase una base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ de V tal que $\alpha_1 = \alpha$, y sea f el funcional lineal que asigna a cada vector en V su primera coordenada en la base ordenada $\mathcal{B}$.

Teorema 17. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F. Para cada vector α de V se define*

$$L_\alpha(f) = f(\alpha), \qquad f \text{ en } V^*.$$

La aplicación $\alpha \to L_\alpha$ es entonces un isomorfismo de V sobre V^{**}.

Demostración. Hemos mostrado que para todo α la función L_α es lineal. Supóngase que α y β pertenezcan a V y c a F, y sea $\gamma = c\alpha + \beta$. Entonces para cada f en V^*

$$\begin{aligned} L_\gamma(f) &= f(\gamma) \\ &= f(c\alpha + \beta) \\ &= cf(\alpha) + f(\beta) \\ &= cL_\alpha(f) + L_\beta(f) \end{aligned}$$

y así

$$L_\gamma = cL_\alpha + L_\beta.$$

Esto muestra que la aplicación $\alpha \rightarrow L_\alpha$ es una transformación lineal de V en V^{**}. Esta transformación es no singular; en efecto, de acuerdo con las observaciones anteriores $L_\alpha = 0$ si, y solo si, $\alpha = 0$. Entonces $\alpha \rightarrow L_\alpha$ es una transformación lineal no singular de V en V^{**}, y como

$$\dim V^{**} = \dim V^* = \dim V$$

el Teorema 9 afirma que esta transformación es inversible y es, por tanto, un isomorfismo de V sobre V^{**}. ∎

Corolario. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F. Si L es un funcional lineal en el espacio dual V^* de V, entonces existe un único vector α de V tal que*

$$L(f) = f(\alpha)$$

para todo f de V^*.

Corolario. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F. Toda base de V^* es dual de alguna base de V.*

Demostración. Sea $\mathcal{B}^* = \{f_1, \ldots, f_n\}$ una base de V^*. Por el Teorema 15, existe una base $\{L_1, \ldots, L_n\}$ de V^{**} tal que

$$L_i(f_j) = \delta_{ij}.$$

Por el corolario anterior, para todo i existe un vector α_i de V tal que

$$L_i(f) = f(\alpha_i)$$

para todo f de V^*; es decir, tal que $L_i = L\alpha_i$. Se sigue inmediatamente que $\{\alpha_1, \ldots, \alpha_n\}$ es una base de V y que $\mathcal{B}^*$ es la dual de esta base. ∎

En vista del Teorema 17 es corriente identificar α con L_α y decir que V «es» el espacio dual del V^*, o que los espacios V, V^* están naturalmente en mutua dualidad. Cada uno es el espacio dual del otro. En el último corolario se tiene una ilustración de lo útil que esto puede ser. He aquí otra ilustración.

Si E es un subconjunto de V^*, entonces el anulador E^0 es (técnicamente) un subconjunto de V^{**}. Si se decide identificar V y V^{**} como en el Teorema 17, entonces E^0 es un subespacio de V, vale decir el conjunto de todos los α de V tales que $f(\alpha) = 0$ para todos los f de E. En un corolario del Teorema 16 hemos

observado que cada subespacio W está determinado por su anulador W^0. ¿Cómo está determinado? La respuesta es que W es el subespacio anulado por todos los f de W^0; esto es, la intersección de los espacios nulos de todos los f en W^0. Con nuestra actual nomenclatura para anuladores, la respuesta puede escribirse en forma muy simple: $W = (W^0)^0$.

Teorema 18. *Si S es cualquier subconjunto de un espacio vectorial de dimensión finita V, entonces $(S^0)^0$ es el subespacio generado por S.*

Demostración. Sea W el subespacio generado por S. Es claro que $W^0 = S^0$. Por tanto, lo que se debe demostrar es que $W = W^{00}$. Ya se ha dado una demostración. He aquí otra. Por el Teorema 16

$$\dim W + \dim W^0 = \dim V$$
$$\dim W^0 + \dim W^{00} = \dim V^*$$

y como $\dim V = \dim V^*$, se tiene

$$\dim W = \dim W^{00}.$$

Como W es un subespacio de W^{00}, se ve que $W = W^{00}$. ∎

Los resultados de esta sección son válidos para espacios vectoriales arbitrarios; sin embargo, las demostraciones requieren el uso del llamado axioma de elección. Queremos evitar enredarnos en una larga discusión de ese axioma, de modo que no se abordarán los anuladores en espacios vectoriales generales. Pero hay dos resultados respecto a funcionales lineales en espacios vectoriales arbitrarios, tan fundamentales, que se darán a continuación.

Sea V un espacio vectorial. Queremos definir hiperespacios en V. A menos que V sea de dimensión finita, no se puede hacer eso considerando la dimensión del hiperespacio. Pero se puede expresar la idea de que un espacio N tiene apenas una dimensión menos que V, de la siguiente forma:

1. N es un subespacio propio de V;
2. si W es un subespacio de V que contiene a N, entonces $W = N$ o $W = V$.

Las condiciones (1) y (2) en conjunto dicen que N es un subespacio propio y que no existe un subespacio propio mayor; vale decir, N es un subespacio propio maximal.

Definición. *Si V es un espacio vectorial, un* **hiperespacio** *en V es un subespacio propio maximal de V.*

Teorema 19. *Si f es un funcional lineal no nulo sobre el espacio vectorial V, entonces el espacio nulo de f es un hiperespacio en V. Recíprocamente, todo hiperespacio de V es el espacio nulo de un (no único) funcional lineal no nulo sobre V.*

Demostración. Sea f un funcional linea no nulo sobre V y N_f su espacio nulo. Sea α un vector de V que no esté en N_f; es decir, un vector tal que $f(\alpha) \neq 0$.

Se demostrará que todo vector de V está en el subespacio generado por N_f y α. Ese subespacio consta de todos los vectores

$$\gamma + c\alpha, \qquad \gamma \text{ en } N_f,\ c \text{ en } F.$$

Sea β en V. Se define

$$c = \frac{f(\beta)}{f(\alpha)}$$

que tiene sentido porque $f(\alpha) \neq 0$. Entonces el vector $\gamma = \beta - c\alpha$ pertenece a N_f, ya que

$$\begin{aligned} f(\gamma) &= f(\beta - c\alpha) \\ &= f(\beta) - cf(\alpha) \\ &= 0. \end{aligned}$$

Con lo que β está en el subespacio generado por N_f y α.

Sea ahora N un hiperespacio en V. Sea α algún vector que no pertenece a N. Como N es un subespacio propio maximal, el subespacio generado por N y α es todo el espacio V. Por tanto, todo vector β de V tiene la forma

$$\beta = \gamma + c\alpha, \qquad \gamma \text{ en } N,\ c \text{ en } F.$$

El vector γ y el escalar c están unívocamente determinados por β. Si se tiene también que

$$\beta = \gamma' + c'\alpha, \qquad \gamma' \text{ en } N,\ c' \text{ en } F$$

entonces

$$(c' - c)\alpha = \gamma - \gamma'.$$

Si $c' - c \neq 0$, entonces α estaría en N; luego $c' = c$ y $\gamma' = \gamma$. Otra manera de formular esta conclusión es: Si β pertenece a V, existe un único escalar c tal que $\beta - c\alpha$ pertenece a N. Sea $g(\beta)$ ese escalar. Es fácil ver que g es un funcional lineal en V y que N es el espacio nulo de g. ∎

Lema. *Si f y g son funcionales lineales en el espacio vectorial V, entonces g es un múltiplo escalar de f si, y solo si, el espacio nulo de g contiene al espacio nulo de f; esto es, si, y solo si, $f(\alpha) = 0$ implica que $g(\alpha) = 0$.*

Demostración. Si $f = 0$ entonces también $g = 0$ y g es trivialmente un múltiplo escalar de f. Supóngase que $f \neq 0$ de modo que el espacio nulo N_f sea un hiperespacio en V. Elíjase algún vector α en V con $f(\alpha) \neq 0$, y sea

$$c = \frac{g(\alpha)}{f(\alpha)}.$$

El funcional lineal $h = g - cf$ es 0 sobre N_f, ya que tanto f como g son ahí 0 y $h(\alpha) = g(\alpha) - cf(\alpha) = 0$. Así h es 0 en el subespacio generado por N_f y α —y ese subespacio es V. Se concluye que $h = 0$; es decir, que $g = cf$.

Teorema 20. *Sean $g, f_1, \ldots, f_r$ funcionales lineales sobre un espacio vectorial V con espacios nulos $N, N_1, \ldots, N_r$, respectivamente. Entonces g es una*

combinación lineal de los $f_1, \ldots, f_r$ *si, y solo si,* N *contiene la intersección* $N_1 \cap \cdots \cap N_r$.

Demostración. Si $g = c_1 f_1 + \cdots + c_r f_r$ y $f_i(\alpha) = 0$ para todo i, entonces evidentemente, $g(\alpha) = 0$. Por tanto, N contiene $N_1 \cap \cdots \cap N_r$.

Demostraremos el recíproco (la parte «si» del teorema) por inducción sobre el número r. El lema precedente se refiere al caso en que $r = 1$. Supóngase que el resultado sea válido para $r = k - 1$ y sean $f_1, \ldots, f_k$ funcionales lineales con espacios nulos $N_1, \ldots, N_k$ tales que $N_1 \cap \cdots \cap N_k$ está contenido en N, el espacio nulo de g. Sean $g', f_1', \ldots, f_{k-1}'$ las restricciones de $g, f_1, \ldots, f_{k-1}$ al subespacio N_k. Entonces $g', f_1', \ldots, f_{k-1}'$ son funcionales lineales sobre el espacio vectorial N_k. Además, si α es un vector de N_k y $f_i'(\alpha) = 0, i = 1, \ldots, k - 1$, entonces α está en $N_1 \cap \cdots \cap N_k$, con lo que $g'(\alpha) = 0$. Por la hipótesis de inducción (el caso $r = k - 1$), existen escalares c_i tales que

$$g' = c_1 f_1' + \cdots + c_{k-1} f_{k-1}'.$$

Sea ahora

(3-16) $$h = g - \sum_{i=1}^{k-1} c_i f_i.$$

Con lo que h es un funcional lineal en V y (3-16) dice que $h(\alpha) = 0$ para todo α en N_k. Por el lema anterior h es un múltiplo escalar de f_k. Si $h = c_k f_k$, entonces

$$g = \sum_{i=1}^{k} c_i f_i. \quad \blacksquare$$

Ejercicios

1. Sea n un entero positivo y F un cuerpo. Sea W el conjunto de todos los vectores $(x_1, \ldots, x_n)$ de F^n tales que $x_1 + \cdots + x_n = 0$.

(a) Demostrar que W^0 consta de todos los funcionales lineales f de la forma

$$f(x_1, \ldots, x_n) = c \sum_{j=1}^{n} x_j.$$

(b) Hacer ver que el espacio dual W^* de W puede identificarse «naturalmente» con los funcionales lineales

$$f(x_1, \ldots, x_n) = c_1 x_1 + \cdots + c_n x_n$$

sobre F^n que satisface $c_1 + \cdots + c_n = 0$.

2. Usando el Teorema 20, demostrar lo siguiente. Si W es un subespacio de un espacio vectorial de dimensión finita V y si $\{g_1, \ldots, g_r\}$ es cualquier base para W^0, entonces

$$W = \bigcap_{i=1}^{r} N_{g_i}.$$

3. Sea S un conjunto, F un cuerpo y $V(S; F)$ el espacio de todas las funciones de S en F:

$$(f + g)(x) = f(x) + g(x)$$
$$(cf)(x) = cf(x).$$

Sea W cualquier subespacio de dimensión n de $V(S; F)$. Demostrar que existen puntos $x_1, \ldots, x_n$ en S y funciones $f_1, \ldots, f_n$ en W tales que $f_i(x_j) = \delta_{ij}$.

3.7. Transpuesta de una transformación lineal

Supóngase que se tienen dos espacios vectoriales V y W sobre el cuerpo F y una transformación lineal T de V en W. Entonces T induce una transformación lineal de W^* en V^*, como sigue. Supóngase que g es un funcional lineal en W, y sea

(3-17) $$f(\alpha) = g(T\alpha)$$

para cada α en V. Entonces (3-17) define una función f de V en F, que es la composición de T, función de V en W, con g, función de W en F. Como ambas, T y g, son lineales, el Teorema 6 dice que f es también lineal; vale decir, f es una función lineal en V. Así T suministra una correspondencia T^t que asocia a cada funcional lineal g sobre W un funcional lineal $f = T^tg$ sobre V, definido por (3-17). Obsérvese también que T^t es igualmente una transformación lineal de W^*; en efecto, si g_1 y g_2 están en W^* y c es un escalar

$$\begin{aligned}[T^t(cg_1 + g_2)](\alpha) &= (cg_1 + g_2)(T\alpha)\\ &= cg_1(T\alpha) + g_2(T\alpha)\\ &= c(T^tg_1)(\alpha) + (T^tg_2)(\alpha)\end{aligned}$$

de modo que $T^t(cg_1 + g_2) = cT^tg_1 + T^tg_2$. Resumamos.

Teorema 21. *Sean V y W espacios vectoriales sobre el cuerpo F. Para toda transformación lineal T de V en W, existe una única transformación lineal T^t de W^* en V^* tal que*

$$(T^tg)(\alpha) = g(T\alpha)$$

para todo g de W^* y todo α de V.

A T^t se la llama **transpuesta** de T. Esta transformación T^t también se llama a menudo adjunta de T, pero no usaremos esta terminología.

Teorema 22. *Sean V y W espacios vectoriales sobre el cuerpo F y sea T una transformación lineal de V en W. El espacio nulo de T^t es el anulador de la imagen de T. Si V y W son de dimensión finita, entonces*

(i) *rango (T^t) = rango (T)*
(ii) *la imagen de T^t es el anulador del espacio nulo de T.*

Demostración. Si g pertenece a W^*, entonces por definición

$$(T^tg)(\alpha) = g(T\alpha)$$

para todo α de V. La afirmación de que g está en el espacio nulo de T^t quiere decir que $g(T\alpha) = 0$ para todo α de V. Así el espacio nulo de T^t es, precisamente, el anulador de la imagen de T.

Supóngase que V y W son de dimensión finita, por ejemplo, $\dim V = n$ y $\dim W = m$. Para (i): Sea r el rango de T, es decir, la dimensión de la imagen de T. Por el Teorema 16 el anulador de la imagen de T tiene entonces dimen-

sión $(m - r)$. Por la primera afirmación de este teorema la nulidad de T^t debe ser $(m - r)$. Pero entonces, como T^t es una transformación lineal en un espacio de dimensión m, el rango de T^t es $m - (m - r) = r$, y así T y T^t tienen el mismo rango. Para (ii): Sea N el espacio nulo de T. Todo funcional en la imagen de T^t está en el anulador de N; en efecto, supóngase que $f = T^t g$ para algún g en W^*; entonces, si α está en N

$$f(\alpha) = (T^t g)(\alpha) = g(T\alpha) = g(0) = 0.$$

Ahora bien, la imagen de T^t es un subespacio del espacio N^0, y

$$\dim N^0 = n - \dim N = \text{rango } (T) = \text{rango } (T^t)$$

con lo que la imagen de T^t debe ser, exactamente, N^0. ∎

Teorema 23. *Sean V y W espacios vectoriales de dimensión finita sobre el cuerpo F. Sea $\mathcal{B}$ una base ordenada de V con base dual $\mathcal{B}^*$, y sea $\mathcal{B}'$ una base ordenada de W con base dual $\mathcal{B}'^*$. Sea T una transformación lineal de V en W; sea A la matriz de T respecto a $\mathcal{B}$, $\mathcal{B}'$ y sea B la matriz de T^t respecto a $\mathcal{B}'^*$, $\mathcal{B}^*$. Entonces $B_{ij} = A_{ji}$.*

Demostración. Sea

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}, \qquad \mathcal{B}' = \{\beta_1, \ldots, \beta_m\},$$
$$\mathcal{B}^* = \{f_1, \ldots, f_n\}, \qquad \mathcal{B}'^* = \{g_1, \ldots, g_m\}.$$

Por definición,

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i, \qquad j = 1, \ldots, n$$

$$T^t g_j = \sum_{i=1}^{n} B_{ij} f_i, \qquad j = 1, \ldots, m.$$

Por otro lado,

$$\begin{aligned}(T^t g_j)(\alpha_i) &= g_j(T\alpha_i) \\ &= g_j\left(\sum_{k=1}^{m} A_{ki}\beta_k\right) \\ &= \sum_{k=1}^{m} A_{ki} g_j(\beta_k) \\ &= \sum_{k=1}^{m} A_{ki}\delta_{jk} \\ &= A_{ji}.\end{aligned}$$

Para cualquier funcional lineal f sobre V

$$f = \sum_{i=1}^{m} f(\alpha_i) f_i.$$

Si se aplica esta fórmula al funcional $f = T^t g_j$ y se considera que $(T^t g_i)(\alpha_i) = A_{ji}$, se tiene

$$T^t g_j = \sum_{i=1}^{n} A_{ji} f_i$$

de donde se desprende en forma inmediata que $B_{ij} = A_{ji}$. ∎

Definición. *Si A es una matriz $m \times n$ sobre el cuerpo F, la* **transpuesta** *de A es la matriz $n \times m$, A^t, definida por $A^t_{ij} = A_{ji}$.*

El Teorema 23 dice, pues, que si T es una transformación lineal de V en W, cuya matriz con respecto a un par de bases es A, entonces la transformación transpuesta T^t está representada, en el par de bases dual, por la matriz transpuesta A^t.

Teorema 24. *Sea A cualquier matriz $m \times n$ sobre el cuerpo F. Entonces el rango de filas de A es igual al rango de columnas de A.*

Demostración. Sea $\mathcal{B}$ la base ordenada canónica de F^n y $\mathcal{B}'$ la base ordenada canónica de F^m. Sea T la transformación lineal de F^n en F^m tal que la matriz de T respecto al par $\mathcal{B}$, $\mathcal{B}'$ sea A; es decir,

$$T(x_1, \ldots, x_n) = (y_1, \ldots, y_m)$$

donde

$$y_i = \sum_{j=1}^{n} A_{ij}x_j.$$

El rango de columnas de A es el rango de la transformación T, pues la imagen de T consta de todos los m-tuples que son combinaciones lineales de los vectores columnas de A.

Respecto a las bases dual $\mathcal{B}'^*$ y $\mathcal{B}^*$, la aplicación transpuesta T^t está representada por la matriz A^t. Como las columnas de A^t son las filas de A, se ve que por la misma razón que el rango de filas de A (el rango de columnas de A^t) es igual al rango de T^t. Por el Teorema 22, T y T^t tienen el mismo rango y, por tanto, el rango de filas de A es igual al rango de columnas de A. ∎

Ahora vemos que si A es una matriz $m \times n$ sobre F y T es la transformación lineal de F^n en F^m definida anteriormente, entonces

$$\text{rango } (T) = \text{rango de filas } (A) = \text{rango de columna } (A)$$

y se dirá simplemente que este número es el **rango** de A.

Ejemplo 25. Este ejemplò será de carácter general; más bien una discusión que un ejemplo. Sea V un espacio vectorial de dimensión n sobre el cuerpo F y sea T un operador lineal sobre V. Supóngase que $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ es una base ordenada de V. La matriz de T en la base ordenada $\mathcal{B}$ está definida como la matriz $n \times n$, A, tal que

$$T\alpha_j = \sum_{j=1}^{n} A_{ij}\alpha_i$$

o sea, que A_{ij} es la i-ésima coordenada del vector $T\alpha_j$ en la base ordenada $\mathcal{B}$. Si $\{f_1, \ldots, f_n\}$ es la base dual de $\mathcal{B}$, esto puede enunciarse simplemente

$$A_{ij} = f_i(T\alpha_j).$$

Se verá ahora qué sucede cuando se cambia de base. Supóngase que

$$\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

es otra base ordenada de V con base dual $\{f'_1, \ldots, f'_n\}$. Si B es la matriz de T en la base ordenada $\mathcal{B}'$, entonces

$$B_{ij} = f'_i(T\alpha'_j).$$

Sea U el operador lineal inversible, tal que $U\alpha_j = \alpha'_j$. Entonces la transpuesta de U viene dada por $U^t f'_i = f_i$. Esto es de fácil verificación, ya que como U es inversible, también lo es U^t y $(U^t)^{-1} = (U^{-1})^t$. Así, $f'_i = (U^{-1})^t f_i$, $i = 1, 2, \ldots, n$. Por tanto,

$$\begin{aligned} B_{ij} &= [(U^{-1})^t f_i](T\alpha'_j) \\ &= f_i(U^{-1}T\alpha'_j) \\ &= f_i(U^{-1}TU\alpha_j). \end{aligned}$$

Y ahora, ¿qué dice esto? Bien, $f_i(U^{-1}TU\alpha_j)$ es el elemento i, j de la matriz $U^{-1}TU$ en la base ordenada $\mathcal{B}$. Los cálculos anteriores indican que este escalar es también el elemento i, j de la matriz de T en la base ordenada $\mathcal{B}'$. Es decir, que

$$\begin{aligned} [T]_{\mathcal{B}'} &= [U^{-1}TU]_{\mathcal{B}} \\ &= [U^{-1}]_{\mathcal{B}}[T]_{\mathcal{B}}[U]_{\mathcal{B}} \\ &= [U]_{\mathcal{B}}^{-1}[T]_{\mathcal{B}}[U]_{\mathcal{B}} \end{aligned}$$

lo que, precisamente, es la fórmula para el cambio de base ya conocida.

Ejercicios

1. Sea F un cuerpo y sea f el funcional lineal en F^2 definido por $f(x_1, x_2) = ax_1 + bx_2$. Para cada uno de los siguientes operadores T, hágase $g = T^t f$ y hállese $g(x_1, x_2)$.

   (a) $T(x_1, x_2) = (x_1, 0)$;
   (b) $T(x_1, x_2) = (-x_2, x_1)$;
   (c) $T(x_1, x_2) = (x_1 - x_2, x_1 + x_2)$.

2. Sea V el espacio vectorial de todas las funciones polinomios sobre el cuerpo de los números reales. Sean a y b dos números reales fijos y sea f el funcional lineal en V definido por

$$f(p) = \int_a^b p(x)\, dx.$$

Si D es el operador derivación sobre V, ¿qué es $D^t f$?

3. Sea V el espacio de las matrices $n \times n$ sobre el cuerpo F y sea B una matriz $n \times n$ dada. Si T es un operador lineal sobre V definido por $T(A) = AB - BA$ y si f es la función traza, ¿qué es $T^t f$?

4. Sea V el espacio vectorial de dimensión finita sobre el cuerpo F y sea T un operador lineal sobre V. Sea c un escalar y supóngase que existe un vector no nulo α en V tal que $T\alpha = c\alpha$. Demostrar que existe un funcional lineal no nulo f en V tal que $T^t f = cf$.

5. Sea A una matriz $m \times n$ de elementos *reales*. Demostrar que $A = 0$ si, y solo si, traza $(A^t A) = 0$.

6. Sea n un entero positivo y sea V el espacio de las funciones polinomios sobre el cuerpo de los números reales que tienen grado n a lo más; es decir, funciones de la forma

$$f(x) = c_0 + c_1x + \cdots + c_nx^n.$$

Sea D el operador derivación sobre V. Hallar una base del espacio nulo del operador transpuesto D^t.

7. Sea V un espacio vectorial de dimensión finita sobre el cuerpo F. Demostrar que $T \rightarrow T^t$ es un isomorfismo de $L(V, V)$ sobre $L(V^*, V^*)$.

8. Sea V el espacio vectorial de las matrices $n \times n$ sobre el cuerpo F.

(a) Si B es una matriz $n \times n$ dada, se define una función f_B sobre V por $f_B(A) =$ traza (B^tA). Demostrar que f_B es un funcional lineal en V.

(b) Demostrar que todo funcional lineal sobre V es de la forma anterior, es decir, es f_B para algún B.

(c) Demostrar que $B \rightarrow f_B$ es un isomorfismo de V sobre V^*.

4. Polinomios

4.1. Algebras

El propósito de este capítulo es establecer algunas de las propiedades básicas del álgebra de polinomios sobre un cuerpo. El tratamiento se facilitará si se introduce primero el concepto de un álgebra lineal sobre un cuerpo.

Definición. *Sea F un cuerpo. Un* **álgebra lineal sobre el cuerpo** *F es un espacio vectorial $\mathfrak{A}$ sobre F con otra operación, llamada* **multiplicación de vectores,** *que asocia a cada par de vectores α, β de $\mathfrak{A}$ un vector $\alpha\beta$ en $\mathfrak{A}$ llamado el producto de α y β, de tal modo que*

(a) *la multiplicación es asociativa,*

$$\alpha(\beta\gamma) = (\alpha\beta)\gamma$$

(b) *la multiplicación es distributiva con respecto a la adición,*

$$\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma \quad \text{y} \quad (\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$$

(c) *para todo escalar c de F,*

$$c(\alpha\beta) = (c\alpha)\beta = \alpha(c\beta).$$

Si existe un elemento 1 en $\mathfrak{A}$ tal que $1\alpha = \alpha 1 = \alpha$ para todo α de $\mathfrak{A}$, $\mathfrak{A}$ se llama u.' **álgebra lineal con unidad sobre** *F, y a 1 se le llama la* **unidad** *de $\mathfrak{A}$. El álgebra $\mathfrak{A}$ se dice* **conmutativa** *si $\alpha\beta = \beta\alpha$, para todo α y β de $\mathfrak{A}$.*

Ejemplo 1. El conjunto de las matrices $n \times n$ sobre un cuerpo, con las operaciones corrientes, es un álgebra lineal con unidad; en particular el cuerpo

mismo es un álgebra lineal con unidad. Este álgebra no es conmutativa si $n \geq 2$. El cuerpo mismo es (evidentemente) conmutativo.

Ejemplo 2. El espacio de todos los operadores lineales sobre un espacio vectorial, con la composición como producto, es un álgebra lineal con unidad. Es conmutativa si, y solo si, el espacio es unidimensional.

Es posible que el lector tenga alguna experiencia con el producto escalar y el producto vectorial de vectores de R^3. De ser así, debe observar que ninguno de estos productos es del tipo descrito en la definición de álgebra lineal. El primer producto es un «producto escalar» en el sentido que asocia a cada par de vectores un escalar, no siendo por consiguiente del tipo de producto que se discute ahora. El producto vectorial asocia un vector a cada par de vectores de R^3; pero ésta no es una multiplicación asociativa.

En lo que resta de esta sección nos ocuparemos de la construcción de un álgebra que es significativamente diferente de las álgebras de los ejemplos anteriores. Sean F un cuerpo y S el conjunto de los enteros no negativos. Por el Ejemplo 3 del Capítulo 2, el conjunto de las funciones de S en F es un espacio vectorial sobre F. Se representará este espacio vectorial por F^∞. Los vectores de F^∞ son, por tanto, sucesiones infinitas $f = (f_0, f_1, f_2, \ldots)$ de escalares f_i de F. Si $g = (g_0, g_1, g_2, \ldots)$ con g_i en F, y a, b escalares de F, $af + bg$ es una sucesión infinita dada por

(4-1) $$af + bg = (af_0 + bg_0, af_1 + bg_1, af_2 + bg_2, \ldots).$$

Definimos un producto en F^∞ asociando a cada par de vectores f y g de F^∞ el vector fg dado por

(4-2) $$(fg)_n = \sum_{i=0}^{n} f_i g_{n-i}, \qquad n = 0, 1, 2, \ldots.$$

Así
$$fg = (f_0g_0, f_0g_1 + f_1g_0, f_0g_2 + f_1g_1 + f_2g_0, \ldots)$$
y como
$$(gf)_n = \sum_{i=0}^{n} g_i f_{n-i} = \sum_{i=0}^{n} f_i g_{n-i} = (fg)_n$$

para $n = 0, 1, 2, \ldots$, se sigue que la multiplicación es conmutativa, $fg = gf$. Si h también pertenece a F^∞, entonces

$$\begin{aligned}
[(fg)h]_n &= \sum_{i=0}^{n} (fg)_i h_{n-i} \\
&= \sum_{i=0}^{n} \left(\sum_{j=0}^{i} f_j g_{i-j} \right) h_{n-i} \\
&= \sum_{i=0}^{n} \sum_{j=0}^{i} f_j g_{i-j} h_{n-i} \\
&= \sum_{j=0}^{n} f_j \sum_{i=0}^{n-j} g_i h_{n-i-j} \\
&= \sum_{j=0}^{n} f_j (gh)_{n-j} = [f(gh)]_n
\end{aligned}$$

para $n = 0, 1, 2, \ldots$, de modo que

$$(fg)h = f(gh). \tag{4-3}$$

Se deja al lector verificar que la multiplicación definida por (4-2) cumple (b) y (c) en la definición de un álgebra lineal y que el vector $1 = (1, 0, 0, \ldots)$ sirve como unidad para F^∞. Entonces F^∞, con las operaciones definidas anteriormente, es un álgebra lineal conmutativa con unidad sobre el cuerpo F.

El vector $(0, 1, 0, \ldots, 0, \ldots)$ juega un papel destacado en lo que sigue y se representará siempre por x. A lo largo de este capítulo nunca se usará x para indicar un elemento del cuerpo F. El producto de x por sí mismo n veces se representará por x^n y se hará que $x^0 = 1$. Entonces

$$x^2 = (0, 0, 1, 0, \ldots), \qquad x^3 = (0, 0, 0, 1, 0, \ldots)$$

y, en general, para todo entero $k \geq 0$, $(x^k)_k = 1$ y $(x^k)_n = 0$ para todo entero no negativo $n \neq k$. Para concluir esta sección observamos que el conjunto formado por $1, x, x^2, \ldots$, es independiente e infinito. Así que el álgebra F^∞ no es de dimensión finita.

El álgebra F^∞ se llama también **álgebra de las series formales de potencias** sobre F. El elemento $f = (f_0, f_1, f_2, \ldots)$ se suele escribir

$$f = \sum_{n=0}^{\infty} f_n x^n. \tag{4-4}$$

Esta notación es muy conveniente para manipular las operaciones algebraicas. Cuando se use, debe recordarse que es puramente formal. No existen «sumas infinitas» en álgebra y la representación en serie de potencias (4-4) no pretende sugerir nada respecto a convergencia, si es que el lector sabe de qué se trata. Usando sucesiones se puede definir rigurosamente un álgebra en que las operaciones se comporten en forma semejante a la adición y multiplicación de series formales de potencias, sin caer en el riesgo de confundirse acerca de cosas tales como sumas infinitas.

4.2. El Algebra de los polinomios

Se está ahora en condiciones de definir un polinomio sobre el cuerpo F.

Definición. *Sea $F[x]$ el subespacio de F^∞ generado por los vectores $1, x, x^2, \ldots$ Un elemento de $F[x]$ se llama* **polinomio sobre** F.

Como $F[x]$ consta de todas las combinaciones lineales (finitas) de x y sus potencias, un vector no nulo f de F^∞ es un polinomio si, y solo si, existe un entero $n \geq 0$ tal que $f_n \neq 0$ y tal que $f_k = 0$ para todos los enteros $k > n$; este entero (cuando existe) es obviamente único y se llama **grado** de f. Se representará el grado de un polinomio f por grd f, y no se asignará grado al polinomio 0. Si f es un polinomio no nulo de grado n se tiene que

$$f = f_0x^0 + f_1x + f_2x^2 + \cdots + f_nx^n, \qquad f_n \neq 0. \tag{4-5}$$

Los escalares $f_0, f_1, \ldots, f_n$ son llamados a veces los **coeficientes** de f, y se dirá que f es un polinomio con coeficientes en F. Se llamarán **polinomios escalares** a los polinomios de la forma cx^0 y frecuentemente se escribirá c en vez de cx^0. Un polinomio no nulo f de grado n tal que $f_n = 1$ se dice que es un polinomio **mónico.**

El lector observará que los polinomios no son la misma clase de objetos que las funciones polinomios sobre F que se han estudiado varias veces. Si F tiene un número infinito de elementos, existe un isomorfismo natural entre $F[x]$ y el álgebra de las funciones polinomios sobre F. Ello se verá en la próxima sección. Verifiquemos ahora que $F[x]$ es un álgebra.

Teorema 1. *Sean f y g polinomios no nulos sobre F. Entonces*

(i) *fg es un polinomio no nulo;*
(ii) *$grd\ (fg) = grd\ f + grd\ g$;*
(iii) *fg es un polinomio mónico si ambos, f y g, son polinomios mónicos;*
(iv) *fg es un polinomio escalar si, y solo si, ambos f y g son polinomios escalares;*
(v) *si $f + g \neq 0$,*

$$grd\ (f + g) \leq máx.\ (grd\ f,\ grd\ g)$$

Demostración. Supóngase que f tiene grado m y que g tiene grado n. Si k es un entero no negativo,

$$(fg)_{m+n+k} = \sum_{i=0}^{m+n+k} f_i g_{m+n+k-i}.$$

Para que $f_i g_{m+n+k-i} \neq 0$ es necesario que $i \leq m$ y $m + n + k - i \leq n$. Luego es necesario que $m + k \leq i \leq m$, lo que implica $k = 0$ e $i = m$. Así,

(4-6) $$(fg)_{m+n} = f_m g_n$$

y

(4-7) $$(fg)_{m+n+k} = 0, \qquad k > 0.$$

Las afirmaciones (i), (ii), (iii) se desprenden inmediatamente de (4-6) y (4-7), mientras que (iv) es una consecuencia de (i) y (ii). Se deja la comprobación de (v) al lector. ∎

Corolario 1. *El conjunto de todos los polinomios sobre el cuerpo F dado dotado de las operaciones* (4-1) *y* (4-2) *es un álgebra lineal conmutativa con unidad sobre F.*

Demostración. Como las operaciones (4-1) y (4-2) son aquellas definidas en el álgebra F^∞ y como $F[x]$ es un subespacio de F^∞, es suficiente demostrar que el producto de dos polinomios es también un polinomio. Ello es trivial cuando uno de los factores es 0, y en los otros casos se deduce de (i). ∎

Corolario 2. *Supóngase que f, g y h son polinomios sobre el cuerpo F tales que $f \neq 0$ y $fg = fh$. Entonces $g = h$.*

Demostración. Como $fg = fh$, $f(g - h) = 0$ y $f \neq 0$, se sigue inmediatamente de (i) que $g - h = 0$. ▌

Otros hechos más se desprenden fácilmente de la demostración del Teorema 1, y de algunos de ellos haremos mención. Supóngase

$$f = \sum_{i=0}^{m} f_i x^i \quad \text{y} \quad g = \sum_{j=0}^{n} g_j x^j.$$

Entonces de (4-7) se tiene

$$fg = \sum_{s=0}^{m+n} \left(\sum_{r=0}^{s} f_r g_{s-r} \right) x^s. \tag{4-8}$$

El lector deberá comprobar, para el caso particular $f = cx^m$, $g = dx^n$ con c, d en F, que (4-8) se reduce a

$$(cx^m)(dx^n) = cdx^{m+n}. \tag{4-9}$$

Ahora, de (4-9) y las leyes distributivas en $F[x]$, se sigue que el producto en (4-8) también está dado por

$$\sum_{i,j} f_i g_j x^{i+j} \tag{4-10}$$

donde la suma se extiende sobre todos los pares de enteros i, j tales que $0 \leq i \leq m$ y $0 \leq j \leq n$.

Definición. *Sea $\mathfrak{A}$ un álgebra lineal con unidad sobre el cuerpo F. Se indicará la unidad de $\mathfrak{A}$ por 1 y se conviene que $\alpha^0 = 1$ para todo α de $\mathfrak{A}$. Entonces a cada polinomio $f = \sum_{i=0}^{n} f_i x^i$ sobre F y α de $\mathfrak{A}$ se asocia un elemento $f(\alpha)$ de $\mathfrak{A}$ por la ley*

$$f(\alpha) = \sum_{i=0}^{n} f_i \alpha^i.$$

Ejemplo 3. Sea C el cuerpo de los números complejos y sea $f = x^2 + 2$.

(a) Si $\mathfrak{A} = C$ y z pertenece a C, $f(z) = z^2 + 2$, en particular $f(2) = 6$ y

$$f\left(\frac{1 + i}{1 - i}\right) = 1.$$

(b) Si $\mathfrak{A}$ es el álgebra de todas las matrices 2×2 sobre C y si

$$B = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$$

entonces

$$f(B) = 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}^2 = \begin{bmatrix} 3 & 0 \\ -3 & 6 \end{bmatrix}.$$

(c) Si $\mathfrak{A}$ es el álgebra de todos los operadores lineales en C^3 y T es el elemento de $\mathfrak{A}$ dado por

$$T(c_1, c_2, c_3) = (i\sqrt{2}\, c_1, c_2, i\sqrt{2}\, c_3)$$

entonces $f(T)$ es el operador lineal sobre C^3 definido por

$$f(T)(c_1, c_2, c_3) = (0, 3c_2, 0).$$

(d) Si $\mathfrak{A}$ es el álgebra de todos los polinomios sobre C y $g = x^4 + 3i$, entonces $f(g)$ es el polinomio de $\mathfrak{A}$ dado por

$$f(g) = -7 + 6ix^4 + x^8.$$

El lector atento habrá observado, en conexión con este último ejemplo, que si f es un polinomio sobre cualquier cuerpo y x es el polinomio $(0, 1, 0, \ldots)$, entonces $f = f(x)$, pero se le aconseja olvidar este hecho.

Teorema 2. *Sea F un cuerpo y $\mathfrak{A}$ un álgebra lineal con unidad sobre F. Supóngase que f y g son polinomios sobre F, que α es un elemento de $\mathfrak{A}$ y que c pertenece a F. Entonces*

(i) $(cf + g)(\alpha) = cf(\alpha) + g(\alpha)$;
(ii) $(fg)(\alpha) = f(\alpha)g(\alpha)$.

Demostración. Como (i) es fácil de probar, se demostrará solamente (ii). Supóngase que

$$f = \sum_{i=0}^{m} f_i x^i \quad \text{y} \quad g = \sum_{j=0}^{n} g_j x^j.$$

Por (4-10),

$$fg = \sum_{i,j} f_i g_j x^{i+j}$$

y luego por (i),

$$\begin{aligned}(fg)(\alpha) &= \sum_{i,j} f_i g_j \alpha^{i+j}\\ &= \left(\sum_{i=0}^{m} f_i \alpha^i\right)\left(\sum_{j=0}^{n} g_j \alpha^j\right)\\ &= f(\alpha)g(\alpha). \quad \blacksquare\end{aligned}$$

Ejercicios

1. Sea F un subcuerpo de los números complejos y sea A la siguiente matriz 2×2 sobre F

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}.$$

Para cada uno de los siguientes polinomios f sobre F, calcular $f(A)$.

(a) $f = x^2 - x + 2$;
(b) $f = x^3 - 1$;
(c) $f = x^2 - 5x + 7$.

2. Sea T el operador lineal T sobre R^3 definido por

$$T(x_1, x_2, x_3) = (x_1, x_3, -2x_2 - x_3).$$

Sea f el polinomio sobre R definido por $f = -x^3 + 2$. Hallar $f(T)$.

3. Sea A una matriz diagonal $n \times n$ sobre el cuerpo F, es decir, una matriz para la cual $A_{ij} = 0$ para $i \neq j$. Sea f el polinomio sobre F definido por

$$f = (x - A_{11}) \cdots (x - A_{nn}).$$

¿Cuál es la matriz $f(A)$?

4. Si f y g son polinomios independientes sobre el cuerpo F y h es un polinomio no nulo sobre F. Demostrar que fh y gh son independientes.

5. Si F es un cuerpo, demostrar que el producto de dos elementos no nulos de F^∞ es no nulo.

6. Sea S un conjunto de polinomios no nulos sobre el cuerpo F. Si no hay dos elementos de S que tengan el mismo grado, demostrar que S es un conjunto independiente de $F[x]$.

7. Si a y b son elementos de un cuerpo F y $a \neq 0$, demostrar que los polinomios $1, ax + b, (ax + b)^2, (ax + b)^3, \ldots$ forman una base de $F[x]$.

8. Si F es un cuerpo y h es un polinomio sobre F de grado ≥ 1, demostrar que la aplicación $f \to f(h)$ es una transformación lineal inyectiva de $F[x]$ en $F[x]$. Demostrar que esta transformación es un isomorfismo de $F[x]$ sobre $F[x]$ si, y solo si, grd $h = 1$.

9. Sea F un subcuerpo de los números complejos y sean T, D las transformaciones sobre $F[x]$ definidas por

$$T\left(\sum_{i=0}^{n} c_i x^i\right) = \sum_{i=0}^{n} \frac{c_i}{1+i} x^{i+1}$$

y

$$D\left(\sum_{i=0}^{n} c_i x^i\right) = \sum_{i=1}^{n} i c_i x^{i-1}.$$

(a) Demostrar que T es un operador lineal no singular sobre $F[x]$. Demostrar también que T no es inversible.

(b) Demostrar que D es un operador lineal sobre $F[x]$ y determinar su espacio nulo.

(c) Demostrar que $DT = I$ y que $TD \neq I$.

(d) Demostrar que $T[(Tf)g] = (Tf)(Tg) - Tf(Tg)$ para todo f, g en $F[x]$.

(e) Formular y demostrar una ley para D análoga a la dada para T en (d).

(f) Supóngase que V es un subespacio no nulo $F[x]$ tal que Tf pertenezca a V para todo f de V. Demostrar que V no es de dimensión finita.

(g) Supóngase que V es un subespacio de dimensión finita de $F[x]$. Demostrar que existe un entero $m \geq 0$ tal que $D^m f = 0$ para todo f en V.

4.3. *Interpolación de Lagrange*

En esta sección F es un cuerpo fijo y $t_0, t_1, \ldots, t_n$ son $n + 1$ elementos *distintos* de F. Sea V el subespacio de $F[x]$ que consta de todos los polinomios de grado menor o igual a n (junto con el polinomio 0), y sea L_i la función de V en F definida para f en V por

$$L_i(f) = f(t_i); \qquad 0 \leq i \leq n.$$

Por la parte (i) del Teorema 2, todo L_i es un funcional lineal sobre V, y una de las cosas que interesan es demostrar que el conjunto constituido por los L_0, $L_1, \ldots, L_n$ es una base de V^*, el espacio dual de V.

Por supuesto, que, para que sea así, es suficiente (en referencia al Teorema 15 del Capítulo 3) que $\{L_0, L_1, \ldots, L_n\}$ sea el dual de una base $\{P_0, P_1, \ldots, P_n\}$ de V. A lo más existe una de tales bases y si existe está caracterizada por

$$L_j(P_i) = P_i(t_j) = \delta_{ij}. \tag{4-11}$$

Los polinomios

$$P_i = \frac{(x - t_0) \cdots (x - t_{i-1})(x - t_{i+1}) \cdots (x - t_n)}{(t_i - t_0) \cdots (t_i - t_{i-1})(t_i - t_{i+1}) \cdots (t_i - t_n)} = \prod_{j \neq i} \left(\frac{x - t_j}{t_i - t_j}\right) \tag{4-12}$$

son de grado n, luego pertenecen a V, y por el Teorema 2 satisfacen (4-11).

Si $f = \sum_i c_i P_i$, entonces para todo j,

$$f(t_j) = \sum_i c_i P_i(t_j) = c_j. \tag{4-13}$$

Como el polinomio 0 tiene la propiedad que $0(t) = 0$ para todo t en F, se sigue de (4-13) que los polinomios $P_0, P_1, \ldots, P_n$ son linealmente independientes. Los polinomios $1, x, \ldots, x^n$ forman una base de V y, por tanto, la dimensión de V es $(n + 1)$. Así el conjunto independiente $\{P_0, P_1, \ldots, P_n\}$ tiene que ser también una base de V. Con lo que para todo f de V

$$f = \sum_{i=0}^{n} f(t_i) P_i. \tag{4-14}$$

La expresión (4-14) se llama la **fórmula de interpolación de Lagrange.** Poniendo $f = x^j$ en (4-14), se tiene

$$x^j = \sum_{i=0}^{n} (t_i)^j P_i.$$

Ahora bien, del Teorema 7 del Capítulo 2, se sigue que la matriz

$$\begin{bmatrix} 1 & t_0 & t_0^2 & \cdots & t_0^n \\ 1 & t_1 & t_1^2 & \cdots & t_1^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^n \end{bmatrix} \tag{4-15}$$

es inversible. La matriz en (4-15) se llama una **matriz de Vandermonde**; es un ejercicio interesante demostrar en forma directa que tal matriz es inversible cuando $t_0, t_1, \ldots, t_n$ son $n + 1$ elementos distintos de F.

Si f es cualquier polinomio sobre F, en el presente tratamiento se representará por $f^\sim$ la función polinomio de F en F que aplica todo t de F en $f(t)$. Por definición (véase Ejemplo 4, Capítulo 2), toda función polinomio surge de esta forma; sin embargo, puede suceder que $f^\sim = g^\sim$ para dos polinomios f y g tales que $f \neq g$. Afortunadamente, como se verá, esta situación poco placentera solo sucede en el caso en que F es un cuerpo que tiene un número finito de elementos distintos. Para describir en forma precisa la relación entre polinomios y funciones polinomios, se necesita definir el producto de dos funcio-

nes polinomios. Si f, g son polinomios sobre F, el producto de $f^\sim$ y $g^\sim$ es la función $f^\sim g^\sim$ de F en F dada por

(4-16) $$(f^\sim g^\sim)(t) = f^\sim(t)g^\sim(t), \qquad t \text{ en } F.$$

Por la parte (ii) del Teorema 2, $(fg)(t) = f(t)g(t)$, y entonces

$$(fg)^\sim(t) = f^\sim(t)g^\sim(t)$$

para cada t en F. Así, $f^\sim g^\sim = (fg)^\sim$, y es una función polinomio. Ahora ya es inmediato, y se deja al lector, verificar que el espacio vectorial de las funciones polinomios sobre F es un álgebra lineal con unidad sobre F si la multiplicación está definida por (4-16).

Definición. *Sea F un cuerpo y sean $\mathfrak{A}$ y $\mathfrak{A}^\sim$ dos álgebras lineales sobre F. Las álgebras $\mathfrak{A}$ y $\mathfrak{A}^\sim$ se dicen isomorfas si existe una aplicación biyectiva $\alpha \rightarrow \alpha^\sim$ de $\mathfrak{A}$ en $\mathfrak{A}^\sim$ tal que*

(a) $$(c\alpha + d\beta)^\sim = c\alpha^\sim + d\beta^\sim$$
(b) $$(\alpha\beta)^\sim = \alpha^\sim\beta^\sim$$

para todo α, β de $\mathfrak{A}$ y todo escalar c, d de F. De la aplicación $\alpha \rightarrow \alpha^\sim$ se dice que es un **isomorfismo de $\mathfrak{A}$ sobre $\mathfrak{A}^\sim$**. *Un isomorfismo de $\mathfrak{A}$ sobre $\mathfrak{A}^\sim$ es entonces un isomorfismo de espacios vectoriales de $\mathfrak{A}$ sobre $\mathfrak{A}^\sim$ que tiene, además, la propiedad (b) de «preservar» productos.*

Ejemplo 4. Sea V un espacio vectorial de dimensión n sobre el cuerpo F. Por el Teorema 13 del Capítulo 3 y las observaciones subsecuentes, cada base ordenada $\mathfrak{B}$ de V determina un isomorfismo $T \rightarrow [T]_{\mathfrak{B}}$ del álgebra de los operadores lineales sobre V sobre el álgebra de las matrices $n \times n$ sobre F. Supóngase ahora que U es un operador lineal fijo y que se ha dado un polinomio

$$f = \sum_{i=0}^{n} c_i x^i$$

con coeficientes c_i en F. Entonces

$$f(U) = \sum_{i=0}^{n} c_i U^i$$

y como $T \rightarrow [T]_{\mathfrak{B}}$ es una aplicación lineal,

$$[f(U)]_{\mathfrak{B}} = \sum_{i=0}^{n} c_i [U^i]_{\mathfrak{B}}.$$

Ahora, del hecho adicional que

$$[T_1T_2]_{\mathfrak{B}} = [T_1]_{\mathfrak{B}}[T_2]_{\mathfrak{B}}$$

para todo T_1, T_2 en $L(V, V)$, se sigue que

$$[U^i]_{\mathfrak{B}} = ([U]_{\mathfrak{B}})^i, \qquad 2 \leq i \leq n.$$

Como esta relación es válida también para $i = 0$, 1 se tiene el resultado

(4-17) $$[f(U)]_{\mathfrak{B}} = f([U]_{\mathfrak{B}}).$$

En otras palabras, si U es un operador lineal sobre V, la matriz de un polinomio en U, en una base dada es el mismo polinomio en la matriz de U.

Teorema 3. *Si F es un cuerpo que tiene un número infinito de elementos distintos, la aplicación $f \to f^{\sim}$ es un isomorfismo del álgebra de polinomios sobre F, sobre el álgebra de las funciones polinomios sobre F.*

Demostración. Por definición, la aplicación es sobreyectiva y si f, g pertenecen a $F[x]$ es evidente que

$$(cf + dg)^{\sim} = df^{\sim} + dg^{\sim}$$

para todo escalar c y d. Como ya se ha visto que $(fg)^{\sim} = f^{\sim}g^{\sim}$, se necesita solo demostrar que la aplicación es inyectiva. Para ello es suficiente, por la linealidad, demostrar que $f^{\sim} = 0$ implica $f = 0$. Supóngase, entonces, que f es un polinomio de grado n, o menor, tal que $f^{\sim} = 0$. Sean $t_0, t_1, \ldots, t_n$, $n + 1$ elementos distintos cualesquiera de F. Como $f^{\sim} = 0$, $f(t_i) = 0$ para $i = 0, 1, \ldots, n$, y es una inmediata consecuencia de (4-14) que $f = 0$. ▮

De los resultados de la siguiente sección obtendremos una demostración totalmente diferente de este teorema.

Ejercicios

1. Usando la fórmula de interpolación de Lagrange, hallar un polinomio f, con coeficientes reales, tal que f tenga grado ≤ 3 y que $f(-1) = -6$, $f(0) = 2$, $f(1) = -2$, $f(2) = 6$.

2. Sean α, β, γ, δ números reales. Se pregunta, ¿cuándo es posible hallar un polinomio f sobre R, de *grado no mayor que* 2, tal que $f(-1) = \alpha$, $f(1) = \beta$, $f(3) = \gamma$ y $f(0) = \delta$? Demostrar que ello solo es posible si, y solo si

$$3\alpha + 6\beta - \gamma - 8\delta = 0.$$

3. Sea F el cuerpo de los números reales,

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$p = (x - 2)(x - 3)(x - 1).$$

(a) Demostrar que $p(A) = 0$.

(b) Sean P_1, P_2, P_3 los polinomios de Lagrange para $t_1 = 2$, $t_2 = 3$, $t_3 = 1$. Calcular $E_i = P_i(A)$, $i = 1, 2, 3$.

(c) Demostrar que $E_1 + E_2 + E_3 = I$, $E_iE_j = 0$ si $i \neq j$, $E_i^2 = E_i$.

(d) Demostrar que $A = 2E_1 + 3E_2 + E_3$.

4. Sea $p = (x - 2)(x - 3)(x - 1)$ y sea T cualquier operador lineal sobre R^4 tal que $p(T) = 0$. Sean P_1, P_2, P_3 los polinomios de Lagrange del Ejercicio 3 y sea $E_i = P_i(T)$, $i = 1, 2, 3$. Demostrar que

$$E_1 + E_2 + E_3 = I, \qquad E_iE_j = 0 \quad \text{si} \quad i \neq j,$$

$$E_i^2 = E_i, \qquad \text{y} \quad T = 2E_1 + 3E_2 + E_3.$$

5. Sea n un entero positivo y F un cuerpo. Supóngase que A es una matriz $n \times n$ sobre F y P una matriz $n \times n$ inversible sobre F. Si f es cualquier polinomio sobre F, demostrar que

$$f(P^{-1}AP) = P^{-1}f(A)P.$$

6. Sea F un cuerpo. Se han considerado ciertos funcionales lineales especiales en $F[x]$ obtenidos por «evaluación en t»:

$$L(f) = f(t).$$

Tales funcionales no son solo lineales, sino que también tienen la propiedad de que $L(fg) = L(f)L(g)$. Demostrar que si L es cualquier funcional lineal sobre $F[x]$ tal que

$$L(fg) = L(f)L(g)$$

para todo f y g, entonces $L = 0$, o existe un t en F tal que $L(f) = f(t)$ para todo f.

4.4. *Ideales de polinomios*

En esta sección se tratarán aquellos temas que dependen ante todo de la estructura multiplicativa del álgebra de los polinomios sobre un cuerpo.

Lema. *Supóngase que f y d son polinomios no nulos sobre un cuerpo F tal que grd $d \leq$ grd f. Entonces existe un polinomio g de $F[x]$ tal que*

$$f - dg = 0 \quad \text{o} \quad grd\,(f - dg) < grd\, f.$$

Demostración. Supóngase que

$$f = a_m x^m + \sum_{i=0}^{m-1} a_i x^i, \qquad a_m \neq 0$$

y que

$$d = b_n x^n + \sum_{i=0}^{n-1} b_i x^i, \qquad b_n \neq 0.$$

Entonces $m \geq n$, y

$$f - \left(\frac{a_m}{b_n}\right)x^{m-n}d = 0 \quad \text{o} \quad \text{grad}\left[f - \left(\frac{a_m}{b_n}\right)x^{m-n}d\right] < \text{grad}\, f.$$

Así que se puede tomar $g = \left(\frac{a_m}{b_n}\right) x^{m-n}$. ∎

Usando este lema se puede ver ahora que el proceso familiar de división de polinomios con coeficientes reales o complejos puede hacerse sobre cualquier cuerpo.

Teorema 4. *Si f, d son polinomios sobre un cuerpo F y d es diferente de* 0, *entonces existen polinomios q, r en $F[x]$ tales que*

(i) $f = dq + r$.
(ii) *o,* $r = 0$ *o* $grd\ r < grd\ d$.

Los polinomios q, r que satisfacen (i) *y* (ii) *son únicos.*

Demostración. Si f es 0 o de grd $f <$ *grd* d, se puede tomar $q = 0$ y $r = f$. En caso de que $f \neq 0$ y grd $f \geq$ grd d, el lema anterior dice que se puede elegir un polinomio g tal que $f - dg = 0$ o grd $(f - dg) <$ grd f. Si $f - dg \neq 0$ y grd $(f - dg) <$ grd d, se elige un polinomio h tal que $(f - dg) - dh = 0$, o

$$\text{grd } [f - d(g + h)] < \text{grd } (f - dg).$$

Si se sigue este proceso tantas veces como sea necesario se llega a obtener polinomios q, r tales que $r = 0$ o grd $r <$ grd d, y $f = dq + r$. Ahora supóngase que también se tenga $f - dq_1 + r_1$, donde $r_1 = 0$ o grd $r_1 <$ grd d. Entontonces $dq + r = dq_1 + r_1$ y $d(q - q_1) = r_1 - r$. Si $q - q_1 \neq 0$, entonces $d(q - q_1) \neq 0$, y

$$\text{grd } d + \text{grd } (q - q_1) = \text{grd } (r_1 - r).$$

Pero como el grado de $r_1 - r$ es menor que el grado de d, esto es imposible y $q - q_1 = 0$. Luego también $r_1 - r = 0$. ▮

Definición. *Sea d un polinomio no nulo sobre el cuerpo F. Si f pertenece a $F[x]$, el teorema anterior dice que existe, a lo más, un polinomio q en $F[x]$ tal que $f = dg$. Si tal q existe, se dice que d* **divide** *a f, que f es* **divisible** *por d, que f es un* **múltiplo** *de d y que q es el* **cociente** *de f por d. Se escribirá, pues, $q = f/d$.*

Corolario 1. *Sea f un polinomio sobre el cuerpo F y sea c un elemento de F. Entonces f es divisible por $x - c$ si, y solo si, $f(c) = 0$.*

Demostración. Por el teorema, $f = (x - c)q + r$, donde r es un polinomio escalar. Por el Teorema 2,

$$f(c) = 0q(c) + r(c) = r(c).$$

Luego $r = 0$ si, y solo si, $f(c) = 0$. ▮

Definición. *Sea F un cuerpo. Un elemento c de F se dice* **raíz,** *o un* **cero,** *de un polinomio dado f sobre F si $f(c) = 0$.*

Corolario 2. *Un polinomio f de grado n sobre un cuerpo F tiene a lo más n raíces en F.*

Demostración. La tesis es obviamente cierta para los polinomios de grado 0 y grado 1. Supóngase que es cierta para polinomios de grado $n - 1$. Si a es una raíz de f, $f = (x - a)q$, donde q tiene grado $n - 1$. Como $f(b) = 0$ si, y solo si, $a = b$ o $q(b) = 0$, se sigue por la hipótesis de inducción que f tiene a lo más n raíces. ▮

El lector debe observar que la etapa principal en la demostración del Teorema 3 se desprende inmediatamente de este corolario.

La derivada formal de un polinomio es de gran utilidad en el estudio de las raíces múltiples. La **derivada** de un polinomio

$$f = c_0 + c_1x + \cdots + c_nx^n$$

es el polinomio

$$f' = c_1 + 2c_2x + \cdots + nc_nx^{n-1}.$$

También se usa la notación $Df = f'$. La derivación es lineal, esto es, D es un operador lineal sobre $F[x]$. Se tienen las derivadas formales de orden superior, $f'' = D^2f$, $f^{(3)} = D^3f$, y así sucesivamente.

Teorema 5 (fórmula de Taylor). *Sea F un cuerpo de característica cero, c un elemento de F y n un entero positivo. Si f es un polinomio sobre f con grd $f \leq n$, entonces*

$$f = \sum_{k=0}^{n} \frac{(D^kf)}{k!}(c)(x - c)^k.$$

Demostración. La fórmula de Taylor es una consecuencia del teorema del binomio y la linealidad de los operadores $D, D^2, \ldots, D^n$. El teorema del binomio es fácilmente demostrable por inducción y dice que

$$(a + b)^m = \sum_{k=0}^{m} \binom{m}{k} a^{m-k} b^k$$

donde

$$\binom{m}{k} = \frac{m!}{k!(m-k)!} = \frac{m(m-1)\cdots(m-k+1)}{1 \cdot 2 \cdots k}$$

son los conocidos coeficientes binomiales que dan el número de combinaciones de m objetos tomados de k en k. Por el teorema del binomio

$$\begin{aligned} x^m &= [c + (x - c)]^m \\ &= \sum_{k=0}^{m} \binom{m}{k} c^{m-k}(x - c)^k \\ &= c^m + mc^{m-1}(x - c) + \cdots + (x - c)^m \end{aligned}$$

que es la fórmula de Taylor para el caso $f = x^m$. Si

$$f = \sum_{m=0}^{n} a_m x^m$$

entonces

$$D^kf(c) = \sum_m a_m(D^kx^m)(c)$$

y

$$\begin{aligned} \sum_{k=0}^{n} \frac{D^kf(c)}{k!}(x - c)^k &= \sum_k \sum_m a_m \frac{(D^kx^m)}{k!}(c)(x - c)^k \\ &= \sum_m a_m \sum_k \frac{(D^kx^m)}{k!}(c)(x - c)^k \\ &= \sum_m a_m x^m \\ &= f. \quad \blacksquare \end{aligned}$$

Debe observarse que, puesto que los polinomios 1, $(x - c)$, . . . , $(x - c)^n$ son linealmente independientes (véase Ejercicio 6, Sección 4.2), la fórmula de Taylor da el único método para escribir f como combinación de los polinomios $(x - c)^k$, $(0 \leq k \leq n)$.

Aunque no se den mayores detalles, es tal vez de utilidad mencionar aquí que, con una apropiada interpretación, la fórmula de Taylor es también válida para polinomios sobre cuerpos de característica finita. Si el cuerpo tiene característica finita (la suma de un número finito de unidades de F es 0), entonces se puede tener que $k! = 0$ en F, en cuyo caso la división de $(D^k f)(c)$ por $k!$ no tiene sentido. No obstante, se puede dar sentido a la división de $D^k f$ por $k!$, porque cada coeficiente de $D^k f$ es un elemento de F multiplicado por un entero divisible por $k!$ Si todo esto resultase confuso, se aconseja al lector restringir su atención a cuerpos de característica 0 o a subcuerpos de los números complejos.

Si c es una raíz del polinomio f, la **multiplicidad** de la raíz c de f es el mayor entero positivo r tal que $(x - c)^r$ divide a f.

La multiplicidad de una raíz es evidentemente menor o igual al grado de f. Para polinomíos sobre cuerpos de característica cero la multiplicidad de la raíz c de f está relacionada con el número de derivadas de f que son nulas en c.

Teorema 6. *Sea F un cuerpo de característica cero y f un polinomio sobre F con grd $f \leq n$. Entonces el escalar c es una raíz de f de multiplicidad r si, y solo si,*

$$(D^k f)(c) = 0, \qquad 0 \leq k \leq r - 1$$
$$(D^r f)(c) \neq 0.$$

Demostración. Supóngase que r es la multiplicidad de la raíz c de f. Entonces existe un polinomio g tal que $f = (x - c)^r g$ y $g(c) \neq 0$. De otro modo, f sería divisible por $(x - c)^{r+1}$, por el Corolario 1 del Teorema 4. Por la fórmula de Taylor aplicada a g

$$f = (x - c)^r \left[\sum_{m=0}^{n-r} \frac{(D^m g)}{m!} (c)\, (x - c)^m \right]$$
$$= \sum_{m=0}^{n-r} \frac{(D^m g)}{m!} (x - c)^{r+m}$$

Como hay solo una manera de escribir f como combinación lineal de las potencias $(x - c)^k$ $(0 \leq k \leq n)$, se sigue que

$$\frac{(D^k f)(c)}{k!} = \begin{cases} 0 \text{ si } 0 \leq k \leq r - 1 \\ \dfrac{D^{k-r} g(c)}{(k - r)!} \text{ si } r \leq k \leq n. \end{cases}$$

Por tanto, $D^k f(c) = 0$ para $0 \leq k \leq x - 1$, y $D^r f(c) = g(c) \neq 0$. Recíproprocamente, si estas condiciones se cumplen, se sigue de inmediato, por la fórmula de Taylor, que existe un polinomio g tal que $f = (x - c)^r g$ y $g(c) \neq 0$. Supóngase ahora que r no sea el mayor entero positivo tal que $(x - c)^r$ divida a f. Entonces existe un polinomio h tal que $f = (x - c)^{r+1} h$. Pero ello implica

$g = (x - c)h$, por el Corolario 2 del Teorema 1; luego $g(c) = 0$, que es una contradicción. ∎

Definición. *Sea F un cuerpo. Un* **ideal** *en $F[x]$ es un subespacio M de $F[x]$ tal que fg pertenece a M siempre que f esté en $F[x]$, y g en M.*

Ejemplo 5. Si F es un cuerpo y d es un polinomio sobre F, el conjunto $M = dF[x]$, de todos los múltiplos df de d por f arbitrario en $F[x]$, es un ideal. En efecto, M no es vacío, ya que contiene a d. Si f, g pertenecen a $F[x]$ y c es un escalar, entonces

$$c(df) - dg = d(cf - g)$$

pertenece a M, con lo que M es un subespacio. Finalmente, también M contiene a $(df)g = d(fg)$. El ideal M se llama el **ideal principal generado** por d.

Ejemplo 6. Sean $d_1, \ldots, d_n$ un número finito de polinomios sobre F. Entonces la suma M de todos los subespacios $d_iF[x]$ es un subespacio y es también un ideal. Para ello supóngase que p pertenezca a M. Entonces existen polinomios $f_1, \ldots, f_n$ en $F[x]$ tal que $p = d_1f_1 + \cdots + d_nf_n$. Si g es un polinomio arbitrario sobre F, entonces

$$pg = d_1(f_1g) + \cdots + d_n(f_ng)$$

de modo que pq también pertenece a M. Así M es un ideal, y se dirá que M es el ideal **generado** por los polinomios $d_1, \ldots, d_n$.

Ejemplo 7. Sea F un subcuerpo de los números reales y considérese el ideal

$$M = (x + 2)F[x] + (x^2 + 8x + 16)F[x].$$

Se afirma que $M = F[x]$. En efecto, M contiene a

$$x^2 + 8x + 16 - x(x + 2) = 6x + 16$$

y, por tanto, M contiene a $6x + 6(x + 2) = 4$. Luego el polinomio escalar 1 pertenece a M, como también todos sus múltiplos.

Teorema 7. *Si F es un cuerpo y M es un ideal no nulo de $F[x]$, existe un único polinomio mónico d en $F[x]$ tal que M es el ideal principal generado por d.*

Demostración. Por suposición, M contiene un polinomio no nulo; entre todos los polinomios no nulos en M existe un polinomio d de menor grado. Podemos admitir que d es un polinomio mónico, pues si no se puede multiplicar d por un escalar para hacerlo mónico. Si f pertenece a M, el Teorema 4 dice que $f = dq + r$, donde $r = 0$ o $\text{grd}\, r < \text{grd}\, d$. Como d está en M, dq y $f - dq = r$ también pertenecen a M. Como d es un elemento de M de grado mínimo, no se puede tener $\text{grd}\, r < \text{grd}\, d$, con lo que $r = 0$. Así $M = dF[x]$. Si g es otro polinomio mónico tal que $M = gF[x]$, entonces existen polinomios no nulos p, q tales que $d = gp$ y $g = dq$. Con lo que $d = dpq$, y

$$\text{grd}\, d = \text{grd}\, d + \text{grd}\, p \quad \text{grd}\, q.$$

Luego grd p = grd $q = 0$, y como d, g son mónicos, $p = q = 1$. Así $d = g$. ▮

Es importante observar que en la demostración recién dada se ha usado el caso especial de un hecho más general y útil; a saber, si p es un polinomio no nulo en un ideal M y si f es un polinomio en M que no es divisible por p, entonces $f = pq + r$, donde el «resto» r que pertenece a M, es distinto de 0, y tiene grado menor que p. Ya se había hecho uso de esta situación en el Ejemplo 7 para ver que el polinomio escalar 1 es el generador mónico del ideal considerado allí. En principio, siempre es posible encontrar el polinomio mónico que genera un ideal dado no nulo. En efecto, se puede obtener, al final, un polinomio en el ideal de grado minimal por un número finito de divisiones sucesivas.

Corolario. *Si $p_1, \ldots, p_n$ son polinomios sobre el cuerpo F, no todos nulos, existe un único polinomio mónico d en $F[x]$ tal que*

(a) *d pertenece al ideal generado por $p_1, \ldots, p_n$;*
(b) *d divide a cada uno de los polinomios p_i.*

Todo polinomio que satisface (a) y (b) necesariamente satisface a

(c) *d es divisible por todo polinomio que divide cada uno de los polinomios $p_1, \ldots, p_n$.*

Demostración. Sea d el generador mónico del ideal

$$p_1F[x] + \cdots + p_nF[x].$$

Cada elemento de este ideal es divisible por d; así, pues, cada uno de los polinomios p_i es divisible por d. Supóngase ahora que f es un polinomio que divide a cada uno de los polinomios $p_1, \ldots, p_n$. Entonces existen polinomios $g_1, \ldots, g_n$ tales que $p_i = fg_i$, $1 \leq i \leq n$. Asimismo, como d pertenece al ideal

$$p_1F[x] + \cdots + p_nF[x],$$

existen polinomios $q_1, \ldots, q_n$ en $F[x]$ tal que

$$d = p_1q_1 + \cdots + p_nq_n.$$

Con lo que

$$d = f[g_1q_1 + \cdots + g_nq_n].$$

Se ha mostrado que d es un polinomio mónico que satisface a (a), (b) y (c). Si d' es cualquier polinomio que satisface (a) y (b) se sigue, de (a) y de la definición de d, que d' es un múltiplo escalar de d y que satisface igualmente a (c). Finalmente, en el caso que d' sea un polinomio mónico, se tiene $d' = d$. ▮

Definición. *Si $p_1, \ldots, p_n$ son polinomios sobre el cuerpo F, no todos nulos, el generador d del ideal*

$$p_1F[x] + \cdots + p_nF[x]$$

se llama el **máximo común divisor** *(m.c.d.) de $p_1, \ldots, p_n$. Esta terminología está justificada por el corolario anterior. Se dice que los polinomios $p_1, \ldots, p_n$ son* **primos relativos** *si su máximo común divisor es 1, o en forma equivalente, si el ideal que ellos generan es todo $F[x]$.*

Ejemplo 8. Sea C el cuerpo de los números complejos. Entonces

(a) m.c.d. $(x + 2, x^2 + 8x + 16) = 1$ (véase Ejemplo 7).

(b) m.c.d. $((x - 2)^2(x + i), (x + 2)(x^2 + 1)) = (x - 2)(x + i)$. En efecto, el ideal

$$(x - 2)^2(x + i)F[x] + (x - 2)(x^2 + 1)F[x]$$

contiene a

$$(x - 2)^2(x + i) - (x - 2)(x^2 + 1) = (x - 2)(x + i)(i - 2).$$

Luego contiene a $(x - 2)(x + i)$, que es mónico y divide a

$$(x - 2)^2(x + i) \quad \text{y} \quad (x - 2)(x^2 + 1).$$

Ejemplo 9. Sea F el cuerpo de los números racionales y en $F[x]$ sea M el ideal generado por

$$(x - 1)(x + 2)^2, \qquad (x + 2)^2(x - 3), \qquad \text{y} \qquad (x - 3).$$

Entonces M contiene a

$$\tfrac{1}{2}(x + 2)^2[(x - 1) - (x - 3)] = (x + 2)^2$$

y como

$$(x + 2)^2 = (x - 3)(x + 7) - 17$$

M contiene al polinomio escalar 1. Así $M = F[x]$ y los polinomios

$$(x - 1)(x + 2)^2, \qquad (x + 2)^2(x - 3), \qquad \text{y} \qquad (x - 3)$$

son primos relativos.

Ejercicios

1. Sea Q el cuerpo de los números racionales. Determinar cuáles de los siguientes subconjuntos de $Q[x]$ son ideales. Cuando el conjunto es un ideal, encontrar su generador mónico.

(a) todos los f de grado par;
(b) todos los f de grado ≥ 5;
(c) todos los f tales que $f(0) = 0$;
(d) todos los f tales que $f(2) = f(4) = 0$;
(e) todos los f en la imagen del operador lineal T definido por

$$T\left(\sum_{i=0}^{n} c_i x^i\right) = \sum_{i=0}^{n} \frac{c_i}{i + 1} x^{i+1}.$$

2. Encontrar el m.c.d. de cada uno de los siguientes pares de polinomios

(a) $2x^5 - x^3 - 3x^2 - 6x + 4, x^4 + x^3 - x^2 - 2x - 2$;
(b) $3x^4 + 8x^2 - 3, x^3 + 2x^2 + 3x + 6$;
(c) $x^4 - 2x^3 - 2x^2 - 2x - 3, x^3 + 6x^2 + 7x + 1$.

3. Sea A una matriz $n \times n$ sobre el cuerpo F. Demostrar que el conjunto de todos los polinomios f en $F[x]$, tales que $f(A) = 0$, es un ideal.

4. Sea F un subcuerpo de los números complejos y sea

$$A = \begin{bmatrix} 1 & -2 \\ 0 & 3 \end{bmatrix}.$$

Encontrar el generador mónico de ideal de todos los polinomios f en $F[x]$ tales que $f(A) = 0$.

5. Sea F un cuerpo. Demostrar que la intersección de cualquier número de ideales en $F[x]$ es un ideal.

6. Sea F un cuerpo. Demostrar que el ideal generado por un número finito de polinomios $f_1, \ldots, f_n$ en $F[x]$ es la intersección de todos los ideales que contienen a $f_1, \ldots, f_n$.

7. Sea K un subcuerpo de un cuerpo F y supóngase que f, g son polinomios en $K[x]$. Sea M_K el ideal generado por f y g en $K[x]$ y M_F el ideal que ellos generan en $F[x]$. Demostrar que M_K y M_F tienen el mismo generador mónico.

4.5. Factorización prima de un polinomio

En esta sección se demostrará que cada polinomio sobre un cuerpo F puede escribirse como producto de polinomios «primos». Esta factorización proporciona un instrumento eficaz para encontrar el máximo común divisor de un número finito de polinomios, y en particular suministra un recurso efectivo para decidir cuándo los polinomios son primos relativos.

Definición. *Sea F un cuerpo. Un polinomio f de $F[x]$ se dice* **reducible sobre** *F si existen polinomios g, h en $F[x]$ de grado ≥ 1 tales que $f = gh$, y si no, se dice que es* **irreducible sobre** *F. Un polinomio no escalar irreducible sobre F se llama* **polinomio primo sobre** *F y a veces se dice solamente que es* **primo en** *$F[x]$.*

Ejemplo 10. El polinomio $x^2 + 1$ es reducible sobre el cuerpo C de los números complejos. En efecto,

$$x^2 + 1 = (x + i)(x - i)$$

y los polinomios $x + i$, $x - i$ pertenecen a $C[x]$. Por otra parte, $x^2 + 1$ es irreducible sobre el cuerpo R de los números reales. Pues si

$$x^2 + 1 = (ax + b)(a'x + b')$$

con a, a', b, b' en R, entonces

$$aa' = 1, \qquad ab' + ba' = 0, \qquad bb' = 1.$$

Estas relaciones implican que $a^2 + b^2 = 0$, lo que es imposible con números reales a y b, a menos que $a = b = 0$.

Teorema 8. *Sean p, f y g polinomios sobre el cuerpo F. Supóngase que p es un polinomio primo y que p divide al producto fg. Entonces p divide a f, o p divide a g.*

Demostración. No se pierde generalidad si se supone que p es un polinomio primo mónico. El hecho de ser p primo simplemente dice que los únicos divi-

sores mónicos de p son 1 y p. Sea d el m.c.d. de f y p. Entonces, $d = 1$ o $d = p$, ya que d es un polinomio mónico que divide a p. Si $d = p$, entonces p divide a f, con lo que se habrá demostrado el teorema. Supóngase entonces que $d = 1$, es decir, supóngase que f y p son primos relativos. Se demostrará que p divide a g. Como $(f, p) = 1$, existen polinomios f_0 y p_0 tales que $1 = f_0 f + p_0 p$. Multiplicando por g, se tiene

$$\begin{aligned} g &= f_0 f g + p_0 p g \\ &= (fg) f_0 + p(p_0 g). \end{aligned}$$

Como p divide a fg, divide también a $(fg)f_0$ y, ciertamente, p divide $p(p_0 g)$. Con lo que p divide a g. ▌

Corolario. *Si p es un primo y divide a un producto $f_1 \cdots f_n$, entonces p divide a uno de los polinomios $f_1, \ldots, f_n$.*

Demostración. La demostración es por inducción. Cuando $n = 2$, la afirmación no es más que el Teorema 6. Supóngase que se ha demostrado el corolario para $n = k$ y que p divide al producto $f_1, \ldots, f_{k+1}$ de ciertos $(k + 1)$ polinomios. Como p divide a $(f_1 \cdots f_k)f_{k+1}$, p divide a f_{k+1}, o p divide a $f_1 \cdots f_k$. Por la hipótesis inductiva, si p divide a $f_1 \cdots f_k$, entonces p divide a f_j para algún j, $1 \leq j \leq k$. Con lo que se ve que en cualquier caso p debe dividir algún f_j, $1 \leq j \leq k + 1$. ▌

Teorema 9. *Si F es un cuerpo, un polinomio mónico no escalar en $F[x]$ puede descomponerse en producto de primos mónicos en $F[x]$ de una manera única, salvo en lo que respecta al orden.*

Demostración. Supóngase que f sea un polinomio mónico no escalar sobre F. Como los polinomios de grado uno son irreducibles, nada hay que demostrar si grd $f = 1$. Supóngase que f tiene grado $n > 1$. Por inducción se supone que el teorema es válido para todo polinomio mónico no escalar de grado menor que n. Si f es irreducible está ya factorizado como un producto de primos mónicos, y en caso contrario $f = gh$, donde g y h son polinomios mónicos no escalares de grado menor que n. Con lo que g y h pueden ser factorizados en producto de primos mónicos en $F[x]$ y, por tanto, también f. Ahora supóngase que

$$f = p_1 \cdots p_m = q_1 \cdots q_n$$

donde $p_1, \ldots, p_m$ y $q_1, \ldots, q_n$ son primos mónicos de $F[x]$. Entonces p_m divide al producto $q_1 \cdots q_n$. Por el corolario anterior, p_m debe dividir a algún q_i. Como q_i y p_m son ambos primos mónicos, ello quiere decir que

(4-16) $$q_i = p_m.$$

Por (4-16) se ve que $m = n = 1$ si $m = 1$ o $n = 1$. En efecto,

$$\operatorname{grad} f = \sum_{i=1}^{m} \operatorname{grad} p_i = \sum_{j=1}^{n} \operatorname{grad} q_j.$$

En este caso no hay más que demostrar, así que se puede suponer que $m > 1$ y $n > 1$. Reordenando los q se puede suponer entonces que $p_m = q_n$ y que

$$p_1 \cdots p_{m-1}p_m = q_1 \cdots q_{n-1}p_m.$$

Ahora por el Corolario 2 del Teorema 1, se sigue que

$$p_1 \cdots p_{m-1} = q_1 \cdots q_{n-1}.$$

Como el polinomio $p_1 \cdots p_{m-1}$ tiene grado menor que n, nuestra hipótesis inductiva es válida y muestra que la sucesión $q_1, \ldots, q_{n-1}$ es a lo más un reordenamiento de la sucesión $p_1, \ldots, p_{m-1}$. Esto, junto con (4-16), muestra que la factorización de f como producto de primos mónicos es única, salvo en lo que respecta al orden de los factores. ∎

En la antedicha factorización de un polinomio no escalar f dado, algunos de los factores mónicos pueden estar repetidos. Si $p_1, p_2, \ldots, p_r$ son los primos mónicos distintos que aparecen en esta factorización de f, entonces

$$f = p_1^{n_1}p_2^{n_2} \cdots p_r^{n_r}, \tag{4-17}$$

siendo el exponente n_i el número de veces que el primo p_i aparece en la factorización. Esta descomposición es también, obviamente, única, y se llama la **descomposición prima** de f. Es de fácil verificación que cada divisor mónico de f tiene la forma

$$p_1^{m_1}p_2^{m_2} \cdots p_r^{m_r}, \qquad 0 \leq m_i \leq n_i. \tag{4-18}$$

De (4-18) se sigue con que el m.c.d. de un número finito de polinomios mónicos no escalares $f_1, \ldots, f_s$ se obtiene por combinación de todos aquellos primos mónicos que aparecen simultáneamente en la factorización de $f_1, \ldots, f_s$. El exponente con que debe tomarse cada primo es el mayor al que la correspondiente potencia prima es factor de cada f_i. Si ninguna potencia prima (no trivial) es factor de cada f_i, los polinomios son primos relativos.

Ejemplo 11. Supóngase que F es un cuerpo y sean a, b, c elementos distintos de F. Entonces los polinomios $x - a$, $x - b$, $x - c$ son primos mónicos distintos en $F[x]$. Si m, n y s son enteros positivos, $(x - c)^s$ es el m.c.d. de los polinomios

$$(x - b)^n(x - c)^s \quad \text{y} \quad (x - a)^m(x - c)^s$$

mientras que los tres polinomios

$$(x - b)^n(x - c)^s, \qquad (x - a)^m(x - c)^s, \qquad (x - a)^m(x - b)^n$$

son primos relativos.

Teorema 10. *Sea f un polinomio mónico no escalar sobre el cuerpo F y sea*

$$f = p_1^{n_1} \cdots p_k^{n_k}$$

la factorización prima de f. Para todo j, $1 \leq j \leq k$, sea

$$f_j = f/p_j^{n_j} = \prod_{i \neq j} p_i^{n_i}.$$

Entonces $f_1, \ldots, f_k$ son primos relativos.

Demostración. Se deja esta demostración (muy fácil) para el lector. Se ha enunciado en detalle este teorema, porque se hará referencia a él más adelante. ∎

Teorema 11. *Sea f un polinomio sobre el cuerpo F con derivada f'. Entonces f es un producto de polinomios irreducibles distintos sobre F si, y solo si, f y f' son primos relativos.*

Demostración. Supóngase que en la factorización prima de f sobre el cuerpo F algún polinomio primo p (no escalar) está repetido. Entonces $f = p^2h$ para algún h en $F[x]$. Entonces

$$f' = p^2h' + 2pp'h$$

y p es también un divisor de f'. Luego f y f' no son primos relativos.

Supóngase ahora que $f = p_1 \cdots p_k$, donde $p_1, \ldots, p_k$ son polinomios irreducibles no escalares distintos sobre F. Sea $f_j = f/p_j$. Entonces

$$f' = p_1'f_1 + p_2'f_2 + \cdots + p_k'f_k.$$

Sea p un polinomio primo que divide a f y a f'. Entonces $p = p_i$ para algún i. Ahora como p_i divide a f_j para $j \neq i$, y como p_i también divide a

$$f' = \sum_{j=1}^{k} p_j'f_j$$

se ve que p_i debe dividir $p_i'f_i$. Por tanto, p_i divide a f_i o a p_i'. Pero p_i no divide a f_i, pues los $p_1, \ldots, p_k$ son distintos. Así p_i divide a p_i'. Ello no es posible, pues p_i' tiene un grado menor que el grado de p_i. Concluimos que ningún primo divide a f y a f', o que f y f' son primos relativos. ∎

Definición. *El cuerpo F se llama* **algebraicamente cerrado** *si todo polinomio primo sobre F tiene grado* 1.

Para decir que F es algebraicamente cerrado, todo polinomio mónico irreducible no escalar sobre F debe ser de la forma $(x - c)$. Ya hemos observado que todo polinomio de éstos es irreducible para cualquier F. En consecuencia, una definición equivalente de un cuerpo algebraicamente cerrado es la de cuerpo F tal que cada polinomio no escalar f en $F[x]$ puede ser expresado en la forma

$$f = c(x - c_1)^{n_1} \cdots (x - c_k)^{n_k}$$

donde c es un escalar, $c_1, \ldots, c_k$ son elementos distintos de F, y $n_1, \ldots, n_k$ son enteros positivos. Otra formulación es que si f es un polinomio no escalar sobre F, entonces existe un elemento c en F tal que $f(c) = 0$.

El cuerpo R de los números reales no es algebraicamente cerrado, pues el polinomio $(x^2 + 1)$ es irreducible sobre R, pero no es de grado 1, o porque no existe número real c tal que $c^2 + 1 = 0$. El llamado teorema fundamental del álgebra dice que el cuerpo C de los números complejos es algebraicamente cerrado. No se demostrará este teorema; sin embargo, se lo usará más adelante

en este libro. La demostración se omite, en parte, por limitación de espacio y, en parte, porque la demostración depende de una propiedad «no algebraica» del sistema de los números reales. Para una posible demostración, el lector interesado puede consultar el libro de Schreier y Sperner en la bibliografía.

El teorema fundamental del álgebra también explica cuáles son las posibilidades de factorización prima de un polinomio con coeficientes reales. Si f es un polinomio con coeficientes reales y c es una raíz compleja de f, entonces la compleja conjugada $\bar{c}$ es también una raíz de f. Por tanto, aquellas raíces complejas que no son reales deben aparecer en pares conjugados y el conjunto entero de raíces tiene la forma $\{t_1, \ldots, t_k, c_1, \bar{c}_1, \ldots, c_r, \bar{c}_r\}$, donde $t_1, \ldots, t_k$ son reales y $c_1, \ldots, c_r$ son números complejos no reales. Con lo que f se descompone en la forma

$$f = c(x - t_1) \cdots (x - t_k)p_1 \cdots p_r$$

donde p_i es el polinomio cuadrático

$$p_i = (x - c_i)(x - \bar{c}_i).$$

Estos polinomios p_i tienen coeficientes reales. Concluimos que todo polinomio irreducible sobre el cuerpo de los números reales tiene grado 1 o 2. Todo polinomio sobre R es el producto de ciertos factores lineales obtenidos de las raíces reales de f y ciertos polinomios cuadráticos irreducibles.

Ejercicios

1. Sea p un polinomio mónico sobre el cuerpo F y sean f y g polinomios primos relativos sobre F. Demostrar que el m.c.d. de pf y pg es p.

2. Suponiendo demostrado el teorema fundamental del álgebra, demostrar lo siguiente. Si f y g son polinomios sobre el cuerpo de los complejos, entonces el m.c.d. $(f, g) = 1$ si, y solo si, f y g no tienen raíces en común.

3. Sea D el operador de la derivación en el espacio de los polinomios sobre el cuerpo de los números complejos. Sea f un polinomio mónico sobre el cuerpo de los números complejos. Demostrar que

$$f = (x - c_1) \cdots (x - c_k)$$

donde $c_1, \ldots, c_k$ son números complejos *distintos* si, y solo si, f y Df son primos relativos. En otras palabras, f no tiene raíces repetidas sino si, y solo si, f y Df no tienen raíces comunes. (Supóngase el teorema fundamental del álgebra.)

4. Demostrar la siguiente generalización de la fórmula de Taylor. Sean f, g y h polinomios sobre un subcuerpo de los números complejos, con grd $f \leq n$. Entonces

$$f(g) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(h)(g - h)^k.$$

(Aquí $f(g)$ representa «f de g».)

Para el resto de los ejercicios se necesita la siguiente definición. Si f, g y p son polinomios sobre el cuerpo F con $p \neq 0$, se dice que f es **congruente con** g **módulo** p si $(f - g)$ es divisible por p. Si f es congruente con g módulo p, se escribe

$$f \equiv g \text{ mód } p.$$

5. Demostrar que para todo polinomio no nulo p, la congruencia módulo p es una relación de equivalencia. Es decir:

(a) Es reflexiva; $f \equiv f$ mód p.
(b) Es simétrica; si $f \equiv g$ mód p, entonces $g \equiv f$ mód p.
(c) Es transitiva; si $f \equiv g$ mód p y $g \equiv h$ mód p, entonces $f \equiv h$ mód p.

6. Supóngase que $f \equiv g$ mód p y $f_1 \equiv g_1$ mód p.

(a) Demostrar que $f + f_1 \equiv g + g_1$ mód p.
(b) Demostrar que $ff_1 \equiv gg_1$ mód p.

7. Usando el Ejercicio 7, demostrar lo que sigue. Si f, g, h y p son polinomios sobre el cuerpo F y $p \neq 0$ y si $f \equiv g$ mód p, entonces $h(f) \equiv h(g)$ mód p.

8. Si p es un polinomio irreducible y $fg \equiv 0$ mód p, demostrar que $f \equiv 0$ mód p o $g \equiv 0$ mód p. Dar un ejemplo que muestre que esto es falso si p no es irreducible.

5. Determinantes

5.1. Anillos conmutativos

En este capítulo demostraremos lo más esencial sobre determinantes de matrices cuadradas. Haremos esto no solo para matrices sobre un cuerpo, sino también para matrices con elementos que son «escalares» de un tipo más general. Hay dos razones para esta generalidad. Primera, en algunos puntos del capítulo siguiente tendremos que usar determinantes de matrices con polinomios como elementos. Segunda, en el tratamiento de los determinantes que presentaremos no interviene el axioma de los cuerpos que garantiza un inverso multiplicativo para cada elemento no nulo. Por estas razones es apropiado desarrollar la teoría de los determinantes de las matrices cuyos elementos pertenecen a un anillo conmutativo con unidad.

Definición. *Un* **anillo** *es un conjunto K, junto con dos operaciones* $(x, y) \rightarrow x + y$ *y* $(x, y) \rightarrow xy$ *que satisfacen:*

(a) *K es un grupo conmutativo para la operación $(x, y) \rightarrow x + y$ (K es grupo aditivo conmutativo);*

(b) $(xy)z = x(yz)$ *(la multiplicación es asociativa);*

(c) $x(y + z) = xy + xz$; $(y + z)x = yx + zx$ *(se cumplen las dos leyes distributivas).*

Si $xy = yx$ para todo x e y de K, se dice que el anillo es **conmutativo.** *Si existe un elemento 1 en K tal que $1x = x1 = x$ para todo x, se dice que K es un* **anillo con unidad,** *y 1 es la* **unidad** *de K.*

Se trata aquí de anillos conmutativos con unidad. Tales anillos pueden ser descritos brevemente como un conjunto K, junto con dos operaciones que

cumplen todos los axiomas de cuerpo dados en el Capítulo 1, excepto posiblemente el axioma (8) y la condición $1 \neq 0$. Así, un cuerpo es un anillo conmutativo con unidad distinta de cero en que a cada x distinto de 0 (cero) le corresponde un elemento x^{-1} tal que $xx^{-1} = 1$. El conjunto de los enteros, con las operaciones corrientes, es un anillo conmutativo con unidad que no es un cuerpo. Otro anillo conmutativo con unidad es el conjunto de todos los polinomios sobre un cuerpo, junto con la adición y multiplicación que se han definido para los polinomios.

Si K es un anillo conmutativo con unidad, se define una matriz $m \times n$ sobre K como una función A del conjunto de los pares (i, j) de enteros, $1 \leq i \leq m$, $1 \leq j \leq n$, en K. Como es usual, se representará una tal matriz por una disposición rectangular que tiene m filas y n columnas. La suma y el producto de matrices sobre K se definen igual que para las matrices sobre un cuerpo

$$(A + B)_{ij} = A_{ij} + B_{ij}$$
$$(AB)_{ij} = \sum_k A_{ik}B_{kj}$$

estando definida la suma cuando A y B tienen el mismo número de filas y el mismo número de columnas, y el producto cuando el número de columnas de A es igual al número de filas de B. Las propiedades algebraicas básicas de estas operaciones son también válidas. Por ejemplo,

$$A(B + C) = AB + AC, \qquad (AB)C = A(BC), \qquad \text{etc.}$$

Como en el caso de los cuerpos, nos referiremos a los elementos de K como escalares. Podemos definir entonces combinaciones lineales de las filas o columnas de la matriz como ya se hizo antes. En general, todo lo que previamente se hizo para las matrices sobre un cuerpo es válido para matrices sobre K, excluyendo aquellos resultados que dependen de la posibilidad de «dividir» en K.

5.2. *Funciones determinantes*

Sea K un anillo conmutativo con unidad. Vamos a asignar a cada matriz $n \times n$ (cuadrada) sobre K un escalar (elemento de K) llamado determinante de la matriz. Es posible definir el determinante de una matriz cuadrada A escribiendo simplemente una fórmula para este determinante en términos de los elementos de A. Se puede entonces deducir las diversas propiedades de los determinantes partiendo de esta fórmula. Sin embargo, tal fórmula es bastante complicada y, para ganar algunas ventajas técnicas, se procederá como sigue. Se definirá una «función determinante» en $K^{n \times n}$ como una función que asigna a cada matriz $n \times n$ un escalar sobre K, función que tiene estas propiedades especiales: es lineal como función de cada una de las filas de la matriz; su valor es 0 sobre toda matriz que tenga dos filas iguales y su valor sobre la matriz identidad $n \times n$ es 1. Se demostrará que tal función existe, y que es única, es decir, que existe exactamente una función así. Cuando demostremos la unicidad, obtendremos una fórmula explícita para el determinante junto con muchas de sus propiedades.

Esta sección se dedicará a la definición de la «función determinante» y a la demostración de que existe al menos una función semejante.

Definición. *Sea K un anillo conmutativo con unidad, n un entero positivo y sea D una función que asigna a cada matriz $n \times n$ sobre K un escalar $D(A)$ en K. Se dice que D es* **n-lineal** *si para cada i, $1 \leq i \leq n$, D es una función lineal de la i-ésima fila cuando las otras $(n-1)$ filas se dejan fijas.*

Esta definición requiere cierta explicación. Si D es una función de $K^{n \times n}$ en K, y si $\alpha_1, \ldots, \alpha_n$ son las filas de la matriz A, se puede escribir también

$$D(A) = D(\alpha_1, \ldots, \alpha_n)$$

esto es, se puede también considerar D como la función de las filas de A. La afirmación de que D es n-lineal quiere decir entonces que

(5-1) $$D(\alpha_1, \ldots, c\alpha_i + \alpha_i', \ldots, \alpha_n) = cD(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) + D(\alpha_1, \ldots, \alpha_i', \ldots, \alpha_n).$$

Si se fijan todas las filas, excepto la fila i, y se considera D como función de la fila i, es a veces conveniente escribir $D(\alpha_i)$ por $D(A)$. Así (5-1) puede abreviarse como

$$D(c\alpha_i + \alpha_i') = cD(\alpha_i) + D(\alpha_i')$$

siempre que quede claro su significado.

Ejemplo 1. Sean $k_1, \ldots, k_n$ enteros positivos, $1 \leq k_i \leq n$, y sea a un elemento de K. Para toda matriz $n \times n$, A, sobre K, se define

(5-2) $$D(A) = aA(1, k_1) \cdots A(n, k_n).$$

Entonces la función D definida por (5-2) es n-lineal. En efecto, si se considera D como función de la fila i, dejando fijas las otras, se puede escribir

$$D(\alpha_i) = A(i, k_i)b$$

donde b es un elemento fijo de K. Sea $\alpha_i = (A_{i1}', \ldots, A_{in}')$. Entonces se tiene

$$\begin{aligned} D(c\alpha_i + \alpha_i') &= [cA(i, k_i) + A'(i, k_i)]b \\ &= cD(\alpha_i) + D(\alpha_i'). \end{aligned}$$

Con lo que D es una función lineal de cada una de las filas de A.

Una función n-lineal particular de este tipo es

$$D(A) = A_{11}A_{22} \cdots A_{nn}.$$

Es decir, el «producto de los elementos de la diagonal» es una función n-lineal sobre $K^{n \times n}$.

Ejemplo 2. Se definen todas las funciones 2-lineales sobre las matrices

2×2 sobre K. Sea D tal función. Si se representan las filas de la matriz 2×2 por ϵ_1, ϵ_2 se tiene

$$D(A) = D(A_{11}\epsilon_1 + A_{12}\epsilon_2, A_{21}\epsilon_1 + A_{22}\epsilon_2).$$

Utilizando que D es 2-lineal, por (5-1), se tiene

$$\begin{aligned} D(A) &= A_{11}D(\epsilon_1, A_{21}\epsilon_1 + A_{22}\epsilon_2) + A_{12}D(\epsilon_2, A_{21}\epsilon_1 + A_{22}\epsilon_2) \\ &= A_{11}A_{21}D(\epsilon_1, \epsilon_1) + A_{11}A_{22}D(\epsilon_1, \epsilon_2) \\ &\qquad + A_{12}A_{21}D(\epsilon_2, \epsilon_1) + A_{12}A_{22}D(\epsilon_2, \epsilon_2). \end{aligned}$$

Con lo que D está completamente determinado por los cuatro escalares

$$D(\epsilon_1, \epsilon_1), \quad D(\epsilon_1, \epsilon_2), \quad D(\epsilon_2, \epsilon_1), \quad \text{y} \quad D(\epsilon_2, \epsilon_2).$$

El lector fácilmente podrá verificar lo que sigue. Si a, b, c, d son cuatro escalares cualesquiera de K y si se define

$$D(A) = A_{11}A_{21}a + A_{11}A_{22}b + A_{12}A_{21}c + A_{12}A_{22}d$$

entonces D es una función 2-lineal de las matrices 2×2 sobre K y

$$\begin{aligned} D(\epsilon_1, \epsilon_1) &= a, & D(\epsilon_1, \epsilon_2) &= b \\ D(\epsilon_2, \epsilon_1) &= c, & D(\epsilon_2, \epsilon_2) &= d. \end{aligned}$$

Lema. *Una combinación lineal de funciones n-lineales es n-lineal.*

Demostración. Basta demostrar que una combinación lineal de dos funciones n-lineales es n-lineal. Sean D y E funciones n-lineales. Si a y b pertenecen a K, la combinación lineal $aD + bE$ está definida por

$$(aD + bE)(A) = aD(A) + bE(A).$$

Luego si fijamos todas las filas, excepto la fila i,

$$\begin{aligned} (aD + bE)(c\alpha_i + \alpha_i') &= aD(c\alpha_i + \alpha_i') + bE(c\alpha_i + \alpha_i') \\ &= acD(\alpha_i) + aD(\alpha_i') + bcE(\alpha_i) + bE(\alpha_i') \\ &= c(aD + bE)(\alpha_i) + (aD + bE)(\alpha_i'). \quad \blacksquare \end{aligned}$$

Si K es un cuerpo y V es el conjunto de las matrices $n \times n$, el lema anterior dice lo siguiente: El conjunto de las funciones n-lineales en V es un subespacio del espacio de todas las funciones de V en K.

Ejemplo 3. Sea D la función definida sobre las matrices 2×2 sobre K por

$$D(A) = A_{11}A_{22} - A_{12}A_{21}. \tag{5-3}$$

Ahora bien, D es la suma de dos funciones del tipo descrito en el Ejemplo 1:

$$\begin{aligned} D &= D_1 + D_2 \\ D_1(A) &= A_{11}A_{22} \\ D_2(A) &= -A_{12}A_{21}. \end{aligned}$$

Por el lema anterior, D es una función 2-lineal. El lector que haya tenido alguna experiencia con los determinantes, no se sorprenderá de este resultado, ya que

reconocerá en (5-3) la definición corriente del determinante de la matriz 2 × 2. Claro que la función D que se ha definido no es una función 2-lineal típica. Tiene varias propiedades especiales. Anotemos algunas de ellas. Primero, si I es la matriz identidad 2 × 2, entonces $D(I) = 1$, es decir, $D(\epsilon_1, \epsilon_2) = 1$. Segundo, si dos filas de A son iguales, entonces

$$D(A) = A_{11}A_{12} - A_{12}A_{11} = 0.$$

Tercero, si A' es la matriz que se obtiene de la matriz 2 × 2, A, intercambiando filas, entonces $D(A') = -D(A)$; en efecto,

$$\begin{aligned} D(A') &= A'_{11}A'_{22} - A'_{12}A'_{21} \\ &= A_{21}A_{12} - A_{22}A_{11} \\ &= -D(A). \end{aligned}$$

Definición. *Sea D una función n-lineal. Se dice que D es* **alternada** *si se cumplen las siguientes dos condiciones:*

(a) *$D(A) = 0$ cuando dos filas de A son iguales.*

(b) *Si A' es una matriz que se obtiene intercambiando dos filas de A, entonces $D(A') = -D(A)$.*

Demostraremos más adelante que toda función n-lineal D que cumple (a), automáticamente cumple (b). Hemos puesto las dos condiciones en la definición n-lineal alternada por razones de conveniencia. El lector probablemente observará que si D satisface (b) y A es una matriz con dos filas iguales, entonces $D(A) = -D(A)$. Es tentador concluir que D satisface también la condición (a). Esto es cierto, por ejemplo, si K es un cuerpo en el que $1 + 1 \neq 0$, pero en general (a) no es consecuencia de (b).

Definición. *Sea K un anillo conmutativo con* **unidad** *y sea n un entero positivo. Supóngase que D es una función de matrices $n \times n$ sobre K en K. Decimos que D es una función determinante si D es n-lineal, alternada, y si $D(I) = 1$.*

Como dijimos antes, mostraremos finalmente que existe exactamente una función determinante sobre matrices $n \times n$ sobre K. Ello se ve fácilmente para las matrices 1 × 1, $A = [a]$, sobre K. La función D dada por $D(A) = a$ es una función determinante y, evidentemente, es la única función determinante de las matrices 1 × 1. Estamos, pues, en condiciones de considerar el caso para $n = 2$. La función

$$D(A) = A_{11}A_{22} - A_{12}A_{21}$$

es una función determinante, como se vio en el Ejemplo 3. Además, la fórmula encontrada en el Ejemplo 2 muestra que D es la única función determinante sobre las matrices 2 × 2. En efecto, se vio que para cualquier función 2-lineal D

$$D(A) = A_{11}A_{21}D(\epsilon_1, \epsilon_1) + A_{11}A_{22}D(\epsilon_1, \epsilon_2) + A_{12}A_{21}D(\epsilon_2, \epsilon_1) + A_{12}A_{22}D(\epsilon_2, \epsilon_2).$$

Si D es alternada, entonces

$$D(\epsilon_1, \epsilon_1) = D(\epsilon_2, \epsilon_2) = 0$$

$$D(\epsilon_2, \epsilon_1) = -D(\epsilon_1, \epsilon_2) = -D(I).$$

Si D satisface también $D(I) = 1$, entonces

$$D(A) = A_{11}A_{22} - A_{12}A_{21}.$$

Ejemplo 4. Sea F un cuerpo y sea D cualquier función 3-lineal y alternada de las matrices 3×3 sobre el anillo de polinomios $F[x]$.

Sea

$$A = \begin{bmatrix} x & 0 & -x^2 \\ 0 & 1 & 0 \\ 1 & 0 & x^3 \end{bmatrix}.$$

Si designamos las filas de la matriz identidad 3×3 por ϵ_1, ϵ_2, ϵ_3, entonces

$$D(A) = D(x\epsilon_1 - x^2\epsilon_3, \epsilon_2, \epsilon_1 + x^3\epsilon_3).$$

Puesto que D es lineal como función de cada fila,

$$\begin{aligned} D(A) &= xD(\epsilon_1, \epsilon_2, \epsilon_1 + x^3\epsilon_3) - x^2D(\epsilon_3, \epsilon_2, \epsilon_1 + x^3\epsilon_3) \\ &= xD(\epsilon_1, \epsilon_2, \epsilon_1) + x^4D(\epsilon_1, \epsilon_2, \epsilon_3) - x^2D(\epsilon_3, \epsilon_2, \epsilon_1) - x^5D(\epsilon_3, \epsilon_2, \epsilon_3). \end{aligned}$$

Puesto que D es alternada, se sigue que

$$D(A) = (x^4 + x^2)D(\epsilon_1, \epsilon_2, \epsilon_3).$$

Lema. *Sea D una función 2-lineal con la propiedad de que $D(A) = 0$ para todas las matrices 2×2, A, sobre K que tienen filas iguales. Entonces D es alternada.*

Demostración. Lo que hay que demostrar es que si A es una matriz 2×2 y si A' se obtiene intercambiando las filas de A, entonces $D(A') = -D(A)$. Si las filas de A son α y β, ello quiere decir que se debe demostrar que $D(\beta, \alpha) = -D(\alpha, \beta)$. Como D es 2-lineal,

$$D(\alpha + \beta, \alpha + \beta) = D(\alpha, \alpha) + D(\alpha, \beta) + D(\beta, \alpha) + D(\beta, \beta).$$

Por la hipótesis, $D(\alpha + \beta, \alpha + \beta) = D(\alpha, \alpha) = D(\beta, \beta) = 0$. Luego

$$0 = D(\alpha, \beta) + D(\beta, \alpha). \quad \blacksquare$$

Lema. *Sea D una función n-lineal de las matrices $n \times n$ sobre K. Supóngase que D tiene la propiedad de que $D(A) = 0$, siempre que dos filas adyacentes de A sean iguales. Entonces D es alternada.*

Demostración. Debe demostrarse que $D(A) = 0$ cuando dos filas cualesquiera de A son iguales, y que $D(A') = -D(A)$ si A' se obtiene por intercambio de dos filas de A. Primero supóngase que A' se obtiene por intercambio de dos filas adyacentes de A. El lector verá que el razonamiento dado en la demostración del lema anterior se extiende al presente caso y da $D(A') = -D(A)$.

Ahora sea B obtenida por intercambio de dos filas i y j de A, con $i < j$. Se puede obtener B de A por medio de una sucesión de intercambios de pares de filas adyacentes. Se comienza intercambiando la fila i con la fila $(i + 1)$ y continuando hasta que las filas queden en el orden

$$\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_j, \alpha_i, \alpha_{j+1}, \ldots, \alpha_n.$$

Esto requiere $k = j - i$ intercambios de filas adyacentes. Se mueve ahora α_j a la posición i haciendo $(k - 1)$ intercambios de filas adyacentes. Se obtiene así B a partir de A por $k + (k - 1) = 2k - 1$ intercambios de filas adyacentes. Con lo que

$$D(B) = (-1)^{2k-1}D(A) = -D(A).$$

Supóngase que A es cualquier matriz $n \times n$ con dos filas iguales, digamos $\alpha_i = \alpha_j$, con $i < j$. Si $j = i + 1$, entonces A tiene dos filas adyacentes iguales y $D(A) = 0$. Si $j > i + 1$, se intercambian α_{i+1} y α_j y la matriz resultante B tiene dos filas adyacentes iguales, entonces $D(B) = 0$. Por otro lado, $D(B) = -D(A)$, luego $D(A) = 0$. ▮

Definición. *Si $n > 1$ y A es una matriz $n \times n$ sobre K, designemos por $A(i|j)$ la matriz $(n - 1) \times (n - 1)$ que se obtiene eliminando la i-ésima fila y la j-ésima columna de A. Si D es una función $(n - 1)$-lineal y A es una matriz $n \times n$, se hace $D_{ij}(A) = D[A(i|j)]$.*

Teorema 1. *Sea $n > 1$ y sea D una función $(n - 1)$-lineal alternada de las matrices $(n - 1) \times (n - 1)$ sobre K. Para todo j, $1 \leq j \leq n$, la función E_j definida por*

$$E_j(A) = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} D_{ij}(A) \tag{5-4}$$

es una función n-lineal alternada de las matrices $n \times n$, A. Si D es una función determinante, también lo es E_j.

Demostración. Si A es una matriz $n \times n$, $D_{ij}(A)$ es independiente de la fila i de A. Como D es $(n - 1)$-lineal, es claro que D_{ij} es lineal como función de cualquier fila, excepto la fila i. Por tanto, $A_{ij}D_{ij}(A)$ es una función n-lineal de A. Una combinación lineal de funciones n-lineal es n-lineal; luego E_j es n-lineal. Para demostrar que E_j es alternada bastará demostrar que $E_j(A) = 0$ siempre que A tenga dos filas iguales. Supóngase que $\alpha_k = \alpha_{k+1}$. Si $i \neq k$ y $i \neq k + 1$, la matriz $A(i|j)$ tiene dos filas iguales, y así $D_{ij}(A) = 0$. Por tanto,

$$E_j(A) = (-1)^{k+j}A_{kj}D_{kj}(A) + (-1)^{k+1+j}A_{(k+1)j}D_{(k+1)j}(A).$$

Como $\alpha_k = \alpha_{k+1}$,

$$A_{kj} = A_{(k+1)j} \quad \text{y} \quad A(k|j) = A(k + 1|j).$$

Entonces es claro que $E_j(A) = 0$.

Supóngase ahora que D es una función determinante. Si $I^{(n)}$ es la matriz

identidad $n \times n$, entonces $I^{(n)}(j|j)$ es la matriz identidad $(n-1) \times (n-1)$, $I^{(n-1)}$. Como $I_{ij}^{(n)} = \delta_{ij}$, se sigue de (5-4) que

$$E_j(I^{(n)}) = D(I^{(n-1)}). \tag{5-5}$$

Ahora, $D(I^{(n-1)}) = 1$, de donde $E_j(I^{(n)}) = 1$ y E_j es una función determinante. ▮

Corolario. *Sea K un anillo conmutativo con unidad, y sea n un entero positivo. Existe al menos una función determinante sobre* $K^{n \times n}$.

Demostración. Se ha demostrado la existencia de una función determinante sobre matrices 1×1 sobre K y también sobre matrices 2×2 sobre K. El Teorema 1 dice explícitamente cómo construir una función determinante sobre matrices $n \times n$, dada tal función sobre matrices $(n-1) \times (n-1)$. El corolario se demuestra por inducción. ▮

Ejemplo 5. Si B es una matriz 2×2 sobre K, sea

$$|B| = B_{11}B_{22} - B_{12}B_{21}.$$

Entonces $|B| = D(B)$, donde D es la función determinante sobre las matrices 2×2. Hemos mostrado que esta función sobre $K^{2\times 2}$ es única. Sea

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

una matriz 3×3 sobre K. Si se define E_1, E_2, E_3 como en (5-4), entonces

$$E_1(A) = A_{11}\begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{21}\begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{vmatrix} + A_{31}\begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{vmatrix} \tag{5-6}$$

$$E_2(A) = -A_{12}\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} + A_{22}\begin{vmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{vmatrix} - A_{32}\begin{vmatrix} A_{11} & A_{13} \\ A_{21} & A_{23} \end{vmatrix} \tag{5-7}$$

$$E_3(A) = A_{13}\begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix} - A_{23}\begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{vmatrix} + A_{33}\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}. \tag{5-8}$$

Se sigue del Teorema 1 que E_1, E_2, E_3 son funciones determinantes. En realidad, como se verá más adelante, $E_1 = E_2 = E_3$, pero esto no se ve, incluso en este caso sencillo. Se podría, sin embargo, comprobar en forma directa desarrollando cada una de las expresiones anteriores. En vez de hacerlo se dan algunos ejemplos concretos.

(a) Sea $K = R[x]$, y

Entonces

$$E_1(A) = (x-1)\begin{vmatrix} x-2 & 1 \\ 0 & x-3 \end{vmatrix} = (x-1)(x-2)(x-3)$$

$$\begin{aligned} E_2(A) &= -x^2\begin{vmatrix} 0 & 1 \\ 0 & x-3 \end{vmatrix} + (x-2)\begin{vmatrix} x-1 & x^3 \\ 0 & x-3 \end{vmatrix} \\ &= (x-1)(x-2)(x-3) \end{aligned}$$

y

$$\begin{aligned} E_3(A) &= x^3\begin{vmatrix} 0 & x-2 \\ 0 & 0 \end{vmatrix} - \begin{vmatrix} x-1 & x^2 \\ 0 & 0 \end{vmatrix} + (x-3)\begin{vmatrix} x-1 & x^2 \\ 0 & x-2 \end{vmatrix} \\ &= (x-1)(x-2)(x-3). \end{aligned}$$

(b) Sea $K = R$ y

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Entonces

$$E_1(A) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

$$E_2(A) = -\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = 1$$

$$E_3(A) = -\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = 1.$$

Ejercicios

1. Cada una de las siguientes expresiones define una función D sobre el conjunto de las matrices 3×3 sobre el cuerpo de los números reales. ¿En qué casos es D una función 3-lineal?

(a) $D(A) = A_{11} + A_{22} + A_{33}$;
(b) $D(A) = (A_{11})^2 + 3A_{11}A_{22}$;
(c) $D(A) = A_{11}A_{12}A_{33}$;
(d) $D(A) = A_{13}A_{22}A_{32} + 5A_{12}A_{22}A_{32}$;
(e) $D(A) = 0$;
(f) $D(A) = 1$.

2. Verificar directamente que las tres funciones E_1, E_2, E_3 definidas por (5-6), (5-7) y (5-8) son idénticas.

3. Sea K un anillo conmutativo con unidad. Si A es una matriz 2×2 sobre K, la **adjunta** de A es la matriz 2×2, adj A, definida por

$$\text{adj } A = \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}.$$

Si det representa la función determinante única de las matrices 2×2 sobre K, demostrar que

(a) $(\text{adj } A)A = A(\text{adj } A) = (\det A)I$;
(b) $\det (\text{adj } A) = \det (A)$;
(c) $\text{adj } (A^t) = (\text{adj } A)^t$.

(A^t denota la transpuesta de A.)

4. Sea A una matriz 2×2 sobre un cuerpo F. Demostrar que A es inversible si, y solo si, $\det A \neq 0$. Si A es inversible, dar una fórmula para A^{-1}.

5. Sea A una matriz 2×2 sobre un cuerpo F y supóngase que $A^2 = 0$. Demostrar que para todo escalar c, $\det (cI - A) = c^2$.

6. Sea K un subcuerpo de los números complejos y n un entero positivo. Sean $j_1, \ldots, j_n$ y $k_1, \ldots, k_n$ enteros positivos no mayores que n. Para una matriz $n \times n$ sobre K se define

$$D(A) = A(j_1, k_1)A(j_2, k_2) \cdots A(j_n, k_n).$$

Demostrar que D es n-lineal si, y solo si, los enteros $j_1, \ldots, j_n$ son distintos.

7. Sea K un anillo conmutativo con unidad. Demostrar que la función determinante sobre las matrices 2×2, A, sobre K es alternada y 2-lineal como función de las columnas de A.

8. Sea K un anillo conmutativo con unidad. Se define una función D sobre las matrices 3×3 sobre K por la regla

$$D(A) = A_{11} \det \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix} - A_{12} \det \begin{bmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{bmatrix} + A_{13} \det \begin{bmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}.$$

Demostrar que D es alternada y 3-lineal como función de las columnas de A.

9. Sea K un anillo conmutativo con unidad y D una función alternada y n-lineal sobre las matrices $n \times n$ sobre K. Demostrar que

(a) $D(A) = 0$ si una de las filas es 0;

(b) $D(B) = D(A)$ si B se obtiene de A por adición de un múltiplo escalar de una fila de A a otra.

10. Sean F un cuerpo, A una matriz 2×3 sobre F y (c_1, c_2, c_3) el vector en F^3 definido por

$$c_1 = \begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{vmatrix}, \quad c_2 = \begin{vmatrix} A_{13} & A_{11} \\ A_{23} & A_{21} \end{vmatrix}, \quad c_3 = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}.$$

Demostrar que

(a) rango $(A) = 2$ si, y solo si, $(c_1, c_2, c_3) \neq 0$;

(b) si A tiene rango 2, entonces (c_1, c_2, c_3) es una base para el espacio de las soluciones del sistema de ecuaciones $AX = 0$.

11. Sea K un anillo conmutativo con unidad y sea D una función alternada 2-lineal sobre las matrices 2×2 sobre K. Demostrar que $D(A) = (\det A)D(I)$ para toda A. Usar este resultado (no son lícitos cálculos con los coeficientes) para demostrar que $\det (AB) = (\det A)(\det B)$ para cualesquiera matrices 2×2, A y B, sobre K.

12. Sea F un cuerpo y D una función sobre las matrices $n \times n$ sobre F (con valores en F). Supóngase que $D(AB) = D(A)D(B)$ para todo A, B. Demostrar que $D(A) = 0$ para toda A, o bien $D(I) = 1$. En este último caso demostrar que $D(A) \neq 0$ si A es inversible.

13. Sea R el cuerpo de los números reales y sea D una función sobre las matrices 2×2 sobre R, con valores en R, tales que $D(AB) = D(A)D(B)$ para cualesquiera A, B. Supóngase que, además,

$$D\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right) \neq D\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

Demostrar que

(a) $D(0) = 0$;

(b) $D(A) = 0$ si $A^2 = 0$;

(c) $D(AB) = -D(A)$ si B se obtiene por intercambio de las filas (o columnas) de A;

(d) $D(A) = 0$ si una fila (o columna) de A es 0;

(e) $D(A) = 0$ si A es singular.

14. Sea A una matriz 2×2 sobre el cuerpo F. Entonces el conjunto de todas las matrices de la forma $f(A)$, donde f es un polinomio sobre F, es un anillo conmutativo con unidad K. Si B es una matriz 2×2 sobre K, el determinante de B es entonces una matriz 2×2 sobre F, de la forma $f(A)$. Supóngase que I es la matriz unidad 2×2 sobre F y que B es la matriz 2×2 sobre K

$$B = \begin{bmatrix} A - A_{11}I & -A_{12}I \\ -A_{21}I & A - A_{22}I \end{bmatrix}.$$

Demostrar que $\det B = f(A)$, donde $f = x^2 - (A_{11} + A_{22})x + \det A$ y también que $f(A) = 0$.

5.3. *Permutaciones y unicidad de los determinantes*

En esta sección demostraremos la unicidad de la función determinante sobre matrices $n \times n$ sobre K. La demostración conducirá de una manera natural a estudiar las permutaciones y algunas de sus propiedades básicas.

Supóngase que D es una función alternada n-lineal sobre las matrices $n \times n$ sobre K. Sea A una matriz $n \times n$ sobre K con filas $\alpha_1, \alpha_2, \ldots, \alpha_n$. Si se representan las filas de la matriz unidad $n \times n$ sobre K por $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$, entonces

$$\alpha_i = \sum_{j=1}^{n} A(i, j)\epsilon_j, \qquad 1 \leq i \leq n. \tag{5-9}$$

Luego

$$\begin{aligned} D(A) &= D\left(\sum_j A(1, j)\epsilon_j, \alpha_2, \ldots, \alpha_n\right) \\ &= \sum_j A(1, j)D(\epsilon_j, \alpha_2, \ldots, \alpha_n). \end{aligned}$$

Si ahora se remplaza α_2 por $\sum_k A(2, k)\epsilon_k$, se ve que

$$D(\epsilon_j, \alpha_2, \ldots, \alpha_n) = \sum_k A(2, k)D(\epsilon_j, \epsilon_k, \ldots, \alpha_n).$$

Así, pues,

$$D(A) = \sum_{j,k} A(1, j)A(2, k)D(\epsilon_j, \epsilon_k, \ldots, \alpha_n).$$

En $D(\epsilon_j, \epsilon_k, \ldots, \alpha_n)$ se remplaza ahora α_3 por $\Sigma A(3, l)\epsilon_l$ y así sucesivamente. Finalmente, se obtiene una expresión complicada, pero teóricamente importante de $D(A)$, a saber,

$$D(A) = \sum_{k_1, k_2, \ldots, k_n} A(1, k_1)A(2, k_2) \cdots A(n, k_n)D(\epsilon_{k_1}, \epsilon_{k_2}, \ldots, \epsilon_{k_n}). \tag{5-10}$$

En (5-10) la suma se extiende sobre todas las sucesiones $(k_1, k_2, \ldots, k_n)$ de enteros positivos no mayores que n. Esto demuestra que D es una suma finita de funciones del tipo descrito en (5-2). Debe observarse que (5-10) es una consecuencia directa de la suposición de que D es n-lineal, y que un caso particular de (5-10) se obtuvo en el Ejemplo 2. Como D es alternada,

$$D(\epsilon_{k_1}, \epsilon_{k_2}, \ldots, \epsilon_{k_n}) = 0$$

siempre que dos de los índices k_i sean iguales. Una sucesión $(k_1, k_2, \ldots, k_n)$ de enteros positivos no mayores que n, con la propiedad de que no haya dos de los k_i iguales, se llama **permutación de grado n.** En (5-10) necesitamos, por tanto, sumar solamente sobre aquellas sucesiones que son permutaciones de grado n.

Como una sucesión finita, o n-tuple, es una función definida sobre los primeros n enteros positivos, una permutación de grado n puede definirse como una función biyectiva del conjunto $\{1, 2, \ldots, n\}$ sobre sí mismo. Una tal función σ corresponde a un n-tuple $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ y es, por tanto, simplemente una regla para ordenar $1, 2, \ldots, n$ en alguna forma bien definida.

Si D es una función alternada n-lineal y A es una matriz $n \times n$ sobre K, se tiene entonces

$$(5\text{-}11) \qquad D(A) = \sum_{\sigma} A(1, \sigma 1) \cdots A(n, \sigma n) D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n})$$

donde la suma se extiende sobre las distintas permutaciones σ de grado n.

A continuación mostraremos que

$$(5\text{-}12) \qquad D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = \pm D(\epsilon_1, \ldots, \epsilon_n)$$

donde el signo $\pm$ depende solo de las permutaciones σ. La razón de ello es como sigue. La sucesión $(\sigma 1, \sigma 2, \ldots, \sigma n)$ se puede obtener de la sucesión $(1, 2, \ldots, n)$ por un número finito de intercambios de los pares de elementos. Por ejemplo, si $\sigma 1 \neq 1$, se puede transponer 1 y $\sigma 1$, obteniéndose $(\sigma 1, \ldots, \sigma 1, \ldots)$. Procediendo así se puede llegar a la sucesión $(\sigma 1, \ldots, \sigma n)$ después de n o menos de tales intercambios de pares. Como D es alternada, el signo de su valor cambia cada vez que se intercambian dos de las filas ϵ_i y ϵ_j. Así, si se pasa de $(1, 2, \ldots, n)$ a $(\sigma 1, \sigma 2, \ldots, \sigma n)$ por medio de m intercambios de pares (i, j), se tendrá

$$D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = (-1)^m D(\epsilon_1, \ldots, \epsilon_n).$$

En particular, si D es una función determinante

$$(5\text{-}13) \qquad D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = (-1)^m$$

donde m depende solo de σ, y no de D. Así, toda función determinante asigna el mismo valor a la matriz de filas $\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}$, y este valor es 1 o -1.

Sea ahora el siguiente hecho básico respecto de las permutaciones. Si σ es una permutación de grado n, se puede pasar de la sucesión $(1, 2, \ldots, n)$ a la sucesión $(\sigma 1, \sigma 2, \ldots, \sigma n)$ por una sucesión de intercambios de pares, y ello puede hacerse de diversas maneras; pero sin que importe cómo se haya hecho, el número de intercambios empleados es siempre par o impar. La permutación

se llama entonces **par** o **impar**, respectivamente. Se define el **signo** de una permutación por

$$\operatorname{sgn}\sigma = \begin{cases} 1, & \text{si } \sigma \text{ es par} \\ -1, & \text{si } \sigma \text{ es impar} \end{cases}$$

representando aquí el símbolo «1» el entero 1.

Se verá más adelante que esta propiedad básica de las permutaciones se puede deducir de lo que ya se sabe sobre funciones determinantes. Supóngase por ahora que ello sea así. Entonces el entero m en (5-13) es siempre par si σ es una permutación par, y es siempre impar si σ es una permutación impar. Para una función alternada n-lineal se tiene entonces que

$$D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = (\operatorname{sgn}\sigma)D(\epsilon_1, \ldots, \epsilon_n)$$

y por (5-11)

$$\text{(5-14)} \qquad D(A) = \left[\sum_{\sigma} (\operatorname{sgn}\sigma)A(1, \sigma 1) \cdots A(n, \sigma n)\right] D(I).$$

Es claro que I representa la matriz identidad $n \times n$.

De (5-14) se ve que existe precisamente una función determinante sobre matrices $n \times n$ sobre K. Si se representa esta función por det, viene dada por

$$\text{(5-15)} \qquad \det(A) = \sum_{\sigma} (\operatorname{sgn}\sigma)A(1, \sigma 1) \cdots A(n, \sigma n)$$

donde la suma se extiende sobre las distintas permutaciones σ de grado n. Todo esto se puede resumir formalmente como sigue.

Teorema 2 *Sea K un anillo conmutativo con unidad y sea n un entero positivo. Existe exactamente una función determinante sobre el conjunto de las matrices $n \times n$ sobre K, y es la función det definida por* (5-15). *Si D es cualquier función alternada n-lineal sobre $K^{n \times n}$, entonces para toda matriz $n \times n$, A,*

$$D(A) = (\det A)D(I).$$

Este es el teorema que habíamos buscado, pero se ha dejado un vacío en la demostración. Este vacío es la demostración de que —para una permutación dada σ, cuando se pasa de $(1, 2, \ldots, n)$ a $(\sigma 1, \sigma 2, \ldots, \sigma n)$ por intercambio de pares— el número de intercambios es siempre par o impar. Este hecho combinatorio básico puede ser demostrado sin referencia alguna a los determinantes. Pero quisiéramos indicar cómo se desprende de la existencia de una función determinante sobre las matrices $n \times n$.

Sea K el anillo de los enteros. Sea D una función determinante sobre las matrices $n \times n$ sobre K. Sea σ una permutación de grado n y supóngase que se pasa de $(1, 2, \ldots, n)$ a $(\sigma 1, \sigma 2, \ldots, \sigma n)$ por m intercambios de los pares (i, j), $i \neq j$. Como se vio en (5-13)

$$(-1)^m = D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n})$$

esto es, el número $(-1)^m$ debe ser el valor de D sobre la matriz de filas $\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}$. Si

$$D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = 1,$$

entonces m debe ser par. Si

$$D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = -1,$$

entonces m debe ser impar.

Como tenemos una fórmula explícita para el determinante de una matriz $n \times n$ y en esta fórmula intervienen las permutaciones de grado n, concluimos esta sección haciendo algunas observaciones más respecto a las permutaciones. Primero obsérvese que hay exactamente $n! = 1 \cdot 2 \cdots n$ permutaciones de grado n. En efecto, si σ es tal permutación, existen n elecciones posibles para $\sigma 1$; cuando esta elección se ha hecho, existen $(n-1)$ elecciones posibles para $\sigma 2$; luego $(n-2)$ elecciones para $\sigma 3$, y así sucesivamente. En consecuencia, se tienen

$$n(n-1)(n-2) \cdots 2 \cdot 1 = n!$$

permutaciones σ. La fórmula (5-15) para det (A) da así el det (A) como suma de $n!$ términos, uno por cada permutación de grado n. Un término dado es un producto

$$A(1, \sigma 1) \cdots A(n, \sigma n)$$

de n elementos de A, un elemento de cada fila y un elemento de cada columna precedido de signo «+» o «−», según que σ sea permutación par o impar.

Cuando las permutaciones se consideran como biyecciones del conjunto $\{1, 2, \ldots, n\}$ sobre sí mismo, se puede definir un producto de permutaciones. El producto de σ y τ será simplemente la función compuesta $\sigma\tau$ definida por

$$(\sigma\tau)(i) = \sigma(\tau(i)).$$

Si ϵ representa la permutación identidad, $\epsilon(i) = i$, entonces cada σ tiene una inversa σ^{-1} tal que

$$\sigma\sigma^{-1} = \sigma^{-1}\sigma = \epsilon.$$

Se pueden resumir estas observaciones diciendo que, respecto de la operación de composición, el conjunto de las permutaciones de grado n es un grupo. Este grupo se llama **grupo simétrico de grado n.**

Desde el punto de vista de productos de permutaciones, la propiedad básica del signo de una permutación es que

$$\text{sgn}(\sigma\tau) = (\text{sgn}\,\sigma)(\text{sgn}\,\tau). \tag{5-16}$$

Es decir, $\sigma\tau$ es una permutación par si σ y τ son ambas pares o ambas impares, mientras que $\sigma\tau$ es impar si una de las dos permutaciones es impar y la otra es par; se puede ver esto por la definición del signo mediante los sucesivos intercambios de los pares (i, j). Puede ser también instructivo indicar cómo sgn $(\sigma\tau)$ = (sgn σ)(sgn τ) se desprende de una propiedad fundamental de los determinantes.

Sea K el anillo de los enteros y sean σ y τ permutaciones de grado n. Sean $\epsilon_1, \ldots, \epsilon_n$ las filas de la matriz identidad $n \times n$ sobre K. Sea A la matriz con filas $\epsilon_{\tau 1}, \ldots, \epsilon_{\tau n}$, y sea B la matriz con filas $\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}$. La i-ésima fila de A tiene

exactamente un elemento no nulo, a saber, el 1 en la columna τ_i. Por lo que es fácil ver que $\epsilon_{\sigma\tau i}$ es la i-ésima fila del producto de matrices AB. Ahora

$$\det(A) = \operatorname{sgn}\tau, \qquad \det(B) = \operatorname{sgn}\sigma, \qquad \text{y} \qquad \det(AB) = \operatorname{sgn}(\sigma\tau).$$

Así se tendrá que $\operatorname{sgn}(\sigma\tau) = (\operatorname{sgn}\sigma)(\operatorname{sgn}\tau)$, una vez demostrado el siguiente teorema.

Teorema 3. *Sea K un anillo conmutativo con unidad y sean A y B matrices $n \times n$ sobre K. Entonces*

$$\det(AB) = (\det A)(\det B).$$

Demostración. Sea B una matriz $n \times n$ dada sobre K, y para cada matriz $n \times n$, A, defínase $D(A) = \det(AB)$. Si se designan las filas de A por $\alpha_1, \ldots, \alpha_n$, entonces

$$D(\alpha_1, \ldots, \alpha_n) = \det(\alpha_1 B, \ldots, \alpha_n B).$$

Aquí $\alpha_j B$ representa la matriz $1 \times n$ que es el producto de la matriz $1 \times n$, α_j, y la matriz $n \times n$, B. Como

$$(c\alpha_i + \alpha_i')B = c\alpha_i B + \alpha_i' B$$

y det es n-lineal; es fácil ver que D es n-lineal. Si $\alpha_i = \alpha_j$, entonces $\alpha_i B = \alpha_j B$, y como det es alternada

$$D(\alpha_1, \ldots, \alpha_n) = 0.$$

Luego D es alternada. Ahora bien, D es una función alternada n-lineal, y por el Teorema 2

$$D(A) = (\det A)D(I).$$

Pero $D(I) = \det(IB) = \det B$, con lo que

$$\det(AB) = D(A) = (\det A)(\det B). \quad \blacksquare$$

Que $\operatorname{sgn}(\sigma\tau) = (\operatorname{sgn}\sigma)(\operatorname{sgn}\tau)$ es solo uno de los muchos corolarios del Teorema 3. Se considerarán algunos de estos corolarios en la siguiente sección.

Ejercicios

1. Si K es un anillo conmutativo con unidad y A es la matriz sobre K dada por

$$A = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$$

Demostrar que $\det A = 0$.

2. Demostrar que el determinante de la matriz de Vandermonde

$$\begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}$$

es $(b - a)(c - a)(c - b)$.

3. Escribir explícitamente las seis permutaciones de grado 3, decir cuáles son impares y cuáles son pares y usar esto para dar la fórmula completa (5-15) del determinante de una matriz 3×3.

4. Sean σ y τ las permutaciones de grado 4 definidas por $\sigma 1 = 2$, $\sigma 2 = 3$, $\sigma 3 = 4$, $\sigma 4 = 1$, $\tau 1 = 3$, $\tau 2 = 1$, $\tau 3 = 2$, $\tau 4 = 4$.
(a) ¿Es σ impar o par? ¿Es τ impar o par?
(b) Hallar $\sigma\tau$ y $\tau\sigma$.

5. Si A es una matriz $n \times n$ inversible sobre un cuerpo, demostrar que $\det A \neq 0$.

6. Sea A una matriz 2×2 sobre un cuerpo. Demostrar que $\det (I + A) = 1 + \det A$ si, y solo si, traza $(A) = 0$.

7. Una matriz $n \times n$, A, se llama **triangular** si $A_{ij} = 0$, siempre que $i > j$ o si $A_{ij} = 0$ siempre que $i < j$. Demostrar que el determinante de una matriz triangular es el producto $A_{11}A_{22} \cdots A_{nn}$ de los elementos de su diagonal.

8. Sea A una matriz 3×3 sobre el cuerpo de los números complejos. Formamos la matriz $xI - A$ con elementos polinomiales, siendo el elemento i, j de esta matriz el polinomio $\delta_{ij}x - A_{ij}$. Si $f = \det (xI - A)$, demostrar que f es un polinomio mónico de grado 3. Si se escribe

$$f = (x - c_1)(x - c_2)(x - c_3)$$

con números complejos c_1, c_2 y c_3, demostrar que

$$c_1 + c_2 + c_3 = \text{traza}\,(A) \quad \text{y} \quad c_1c_2c_3 = \det A.$$

9. Sea n un entero positivo y F un cuerpo. Si σ es una permutación de grado n, demostrar que la función

$$T(x_1, \ldots, x_n) = (x_{\sigma 1}, \ldots, x_{\sigma n})$$

es un operador lineal inversible en F^n.

10. Sean F un cuerpo, n un entero positivo y S el conjunto de las matrices $n \times n$ sobre F. Sea V el espacio vectorial de todas las funciones de S en F. Sea W el conjunto de las funciones alternadas n-lineales sobre S. Demostrar que W es un subespacio de V. ¿Cuál es la dimensión de W?

11. Sea T un operador lineal sobre F^n. Defínase

$$D_T(\alpha_1, \ldots, \alpha_n) = \det (T\alpha_1, \ldots, T\alpha_n).$$

(a) Demostrar que D_T es una función alternada n-lineal.
(b) Si

$$c = \det (T\epsilon_1, \ldots, T\epsilon_n)$$

demostrar que para n vectores $\alpha_1, \ldots, \alpha_n$ arbitrarios se tiene

$$\det (T\alpha_1, \ldots, T\alpha_n) = c \det (\alpha_1, \ldots, \alpha_n).$$

(c) Si $\mathcal{B}$ es cualquier base ordenada de F^n y A es la matriz de T en la base ordenada $\mathcal{B}$, demostrar que $\det A = c$.
(d) ¿Qué nombre apropiado se podría dar a c?

12. Si σ es una permutación de grado n y A una matriz $n \times n$ sobre el cuerpo F con vectores fila $\alpha_1, \ldots, \alpha_n$, sea $\sigma(A)$ la matriz $n \times n$ con vectores fila $\alpha_{\sigma 1}, \ldots, \alpha_{\sigma n}$.

(a) Demostrar que $\sigma(AB) = \sigma(A)B$ y, en particular, que $\sigma(A) = \sigma(I)A$.

(b) Si T es el operador lineal del Ejercicio 9, demostrar que la matriz de T en la base canónica es $\sigma(I)$.

(c) ¿Es $\sigma^{-1}(I)$ la matriz inversa de $\sigma(I)$?

(d) ¿Es cierto que $\sigma(A)$ es semejante a A?

13. Demostrar que la función signo de las permutaciones es única en el siguiente sentido: Si f es cualquier función que asigna a cada permutación de grado n un entero y si $f(\sigma\tau) = f(\sigma)f(\tau)$, entonces f es idénticamente 0 o f es idénticamente 1 o f es la función signo.

5.4. Otras propiedades de los determinantes

En esta sección mencionaremos algunas de las más importantes propiedades de la función determinante sobre las matrices $n \times n$. Probablemente lo primero que deberíamos señalar es lo siguiente. En el estudio de det A, las filas de A han jugado un papel privilegiado. Como no existe diferencia fundamental entre filas y columnas, se puede muy bien esperar que det A sea una función alternada n-lineal de las columnas de A. Este es el caso, y para demostrarlo es suficiente hacer ver que

$$\det(A^t) = \det(A) \tag{5-17}$$

donde A^t representa la transpuesta de A.

Si σ es una permutación de grado n,

$$A^t(i, \sigma i) = A(\sigma i, i).$$

Por la expresión (5-15) se tiene entonces

$$\det(A^t) = \sum_{\sigma} (\operatorname{sgn} \sigma) A(\sigma 1, 1) \cdots A(\sigma n, n).$$

Si $i = \sigma^{-1}j$, $A(\sigma i, i) = A(j, \sigma^{-1}j)$. Con lo que

$$A(\sigma 1, 1) \cdots A(\sigma n, n) = A(1, \sigma^{-1}1) \cdots A(n, \sigma^{-1}n).$$

Como $\sigma\sigma^{-1}$ es la permutación identidad

$$(\operatorname{sgn} \sigma)(\operatorname{sgn} \sigma^{-1}) = 1 \quad \text{o} \quad \operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma).$$

Además, como σ varía sobre todas las permutaciones de grado n, también lo hace σ^{-1}. Por tanto,

$$\begin{aligned}\det(A^t) &= \sum_{\sigma} (\operatorname{sgn} \sigma^{-1}) A(1, \sigma^{-1}1) \cdots A(n, \sigma^{-1}n) \\ &= \det A\end{aligned}$$

que demuestra (5-17).

En ciertas ocasiones se necesita calcular determinantes dados. Cuando ello es necesario, suele ser útil aprovechar la siguiente propiedad. *Si B se obtiene de A por la adición de un múltiplo de una fila de A a otra (o un múltiplo de una columna a otra), entonces*

$$\det B = \det A. \tag{5-18}$$

Se demostrará la afirmación para las filas. Sea B la que se obtiene de A por la adición de $c\alpha_j$ a α_i, donde $i < j$. Como det es lineal como función de la i-ésima fila

$$\begin{aligned} \det B &= \det A + c \det(\alpha_1, \ldots, \alpha_j, \ldots, \alpha_j, \ldots, \alpha_n) \\ &= \det A. \end{aligned}$$

Otra propiedad útil es la siguiente. Supóngase que tenemos una matriz $n \times n$ en forma de bloque

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

donde A es una matriz $r \times r$, C una matriz $s \times s$, B una matriz $r \times s$ y 0 representa la matriz nula $s \times r$. Entonces

(5-19) $$\det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = (\det A)(\det C).$$

Para demostrarlo, se define

$$D(A, B, C) = \det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}.$$

Si fijamos A y B, entonces D es alternada y s-lineal como función de las filas de C. Así, por el Teorema 2

$$D(A, B, C) = (\det C)D(A, B, I)$$

donde I es la matriz identidad $s \times s$. Restando múltiplos de las filas de I de las filas de B y usando la afirmación (5-18), se tiene

$$D(A, B, I) = D(A, 0, I).$$

Ahora, $D(A, 0, I)$ es evidentemente alternada y r-lineal como función de las filas de A. Así, pues,

$$D(A, 0, I) = (\det A)D(I, 0, I).$$

Pero $D(I, 0, I) = 1$, con lo que

$$\begin{aligned} D(A, B, C) &= (\det C)D(A, B, I) \\ &= (\det C)D(A, 0, I) \\ &= (\det C)(\det A). \end{aligned}$$

Por un razonamiento del mismo tipo, o tomando transpuestas

(5-20) $$\det \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} = (\det A)(\det C).$$

Ejemplo 6. Supóngase que K es el cuerpo de los números racionales y que se desea calcular el determinante de la matriz 4×4

$$A = \begin{bmatrix} 1 & -1 & 2 & 3 \\ 2 & 2 & 0 & 2 \\ 4 & 1 & -1 & -1 \\ 1 & 2 & 3 & 0 \end{bmatrix}.$$

Restando múltiplos adecuados de la fila 1, de las filas 2, 3 y 4, obtenemos la matriz

$$\begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 5 & -9 & -13 \\ 0 & 3 & 1 & -3 \end{bmatrix}$$

que según (5-18) tiene la misma matriz que A. Si se resta $\frac{5}{4}$ veces la fila 2 de la fila 3 y luego se resta $\frac{3}{4}$ veces la fila 2 de la fila 4, se tiene

$$B = \begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 0 & -4 & -8 \\ 0 & 0 & 4 & 0 \end{bmatrix}.$$

y nuevamente $\det B = \det A$. La forma bloque de B dice que

$$\det A = \det B = \begin{vmatrix} 1 & -1 \\ 0 & 4 \end{vmatrix} \begin{vmatrix} -4 & -8 \\ 4 & 0 \end{vmatrix} = 4(32) = 128.$$

Sea ahora $n > 1$ y sea A una matriz $n \times n$ sobre K. En el Teorema 1 se vio cómo construir una función determinante sobre las matrices $n \times n$, dada una de las matrices $(n-1) \times (n-1)$. Ahora que se ha demostrado la unicidad de la función determinante, la fórmula (5-4) dice lo siguiente. Si se fija cualquier columna de índice j,

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det A(i|j).$$

El escalar $(-1)^{i+j} \det A(i|j)$ se suele llamar **cofactor** i, j de A, o cofactor del elemento i, j de A. La expresión anterior para $\det A$ es entonces llamada desarrollo del $\det A$ por cofactores de la j-ésima columna (o, a veces, desarrollo por menores de la j-ésima columna). Si se hace

$$C_{ij} = (-1)^{i+j} \det A(i|j)$$

entonces la expresión anterior dice que para cada j

$$\det A = \sum_{i=1}^{n} A_{ij} C_{ij}$$

donde el cofactor C_{ij} es $(-1)^{i+j}$ veces el determinante de la matriz $(n-1) \times (n-1)$ que se obtiene suprimiendo la fila i y la columna j de A.

Si $j \neq k$, entonces

$$\sum_{i=1}^{n} A_{ik} C_{ij} = 0.$$

En efecto, remplazando la columna j de A por su columna k y llamando B la matriz resultante, entonces B tiene dos columnas iguales y, por tanto, $\det B = 0$. Como $B(i|j) = A(i|j)$, se tiene

$$\begin{aligned}0 &= \det B\\ &= \sum_{i=1}^{n} (-1)^{i+j} B_{ij} \det B(i|j)\\ &= \sum_{i=1}^{n} (-1)^{i+j} A_{ik} \det A(i|j)\\ &= \sum_{i=1}^{n} A_{ik} C_{ij}.\end{aligned}$$

Estas propiedades de los cofactores pueden resumirse con

(5-21) $$\sum_{i=1}^{n} A_{ik} C_{ij} = \delta_{jk} \det A.$$

La matriz $n \times n$, adj A, transpuesta de la matriz de los cofactores de A, se llama **adjunta** de A. Así

(5-22) $$(\text{adj } A)_{ij} = C_{ji} = (-1)^{i+j} \det A(j|i).$$

Las expresiones (5-21) pueden resumirse en la ecuación matricial

(5-23) $$(\text{adj } A)A = (\det A)I.$$

Queremos ver también que $A(\text{adj } A) = (\det A)I$. Como $A^t(i|j) = A(j|i)^t$, se tiene

$$(-1)^{i+j} \det A^t(i|j) = (-1)^{j+i} \det A(j|i)$$

que dice simplemente que el cofactor i, j de A^t es el cofactor j, i de A. Con lo que

(5-24) $$\text{adj } (A^t) = (\text{adj } A)^t$$

Aplicando (5-23) a A^t, se tiene

$$(\text{adj } A^t)A^t = (\det A^t)I = (\det A)I$$

y transponiendo

$$A(\text{adj } A^t)^t = (\det A)I.$$

Usando (5-24) se tiene lo que se deseaba:

(5-25) $$A(\text{adj } A) = (\det A)I.$$

Como para las matrices sobre un cuerpo, una matriz $n \times n$, A, sobre K, se dice **inversible sobre** K si existe una matriz $n \times n$, A^{-1}, con elementos en K, tal que $AA^{-1} = A^{-1}A = I$. Si tal matriz existe, es única; en efecto, con el mismo razonamiento usado en el Capítulo 1, se ve que cuando $BA = AC = I$ tenemos que $B = C$. Las fórmulas (5-23) y (5-25) nos dicen lo siguiente acerca de la inversión de matrices sobre K. Si el elemento det A tiene un inverso multiplicativo en K, entonces A es inversible y $A^{-1} = (\det A)^{-1}$ adj A es la inversa única de A. Recíprocamente, es fácil ver que si A es inversible sobre K, el elemento det A es inversible en K. Para ello, si $BA = I$ se tiene

$$1 = \det I = \det (AB) = (\det A)(\det B).$$

Lo demostrado es el siguiente teorema.

Teorema 4. *Sea A una matriz $n \times n$ sobre K. Entonces A es inversible sobre K si, y solo si, det A es inversible en K. Cuando A es inversible, la inversa única de A es*

$$A^{-1} = (\det A)^{-1} \operatorname{adj} A.$$

En particular, una matriz $n \times n$ sobre un cuerpo es inversible si, y solo si, su determinante es distinto de cero.

Debemos señalar que este criterio decisivo para la inversión demuestra que una matriz $n \times n$ con una inversa a la izquierda o a la derecha es inversible. Esta demostración es completamente independiente de la demostración que se dio en el Capítulo 1 para las matrices sobre un cuerpo. Queremos también indicar qué quiere decir la inversión para matrices con elementos polinomios. Si K es el anillo de los polinomios $F[x]$, los únicos elementos de K que son inversibles son los polinomios escalares no nulos. En efecto, si f y g son polinomios y $fg = 1$, se tiene grd f + grd $g = 0$, con lo que grd f = grd $g = 0$; es decir, f y g son polinomios escalares. Así, una matriz $n \times n$ sobre el anillo de los polinomios $F[x]$ es inversible sobre $F[x]$ si, y solo si, su determinante es un polinomio escalar no nulo.

Ejemplo 7. Sea $K = R[x]$ el anillo de los polinomios sobre el cuerpo de los números reales. Sea

$$A = \begin{bmatrix} x^2 + x & x + 1 \\ x - 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} x^2 - 1 & x + 2 \\ x^2 - 2x + 3 & x \end{bmatrix}.$$

Entonces, por un cálculo breve, se tiene que det $A = x + 1$ y det $B = -6$. Con lo que A no es inversible sobre K. Obsérvese que

$$\operatorname{adj} A = \begin{bmatrix} 1 & -x - 1 \\ -x + 1 & x^2 + x \end{bmatrix}, \qquad \operatorname{adj} B = \begin{bmatrix} x & -x - 2 \\ -x^2 + 2x - 3 & x^2 - 1 \end{bmatrix}$$

y $(\operatorname{adj} A)A = (x + 1)I$, $(\operatorname{adj} B)B = -6I$. Es claro que

$$B^{-1} = -\frac{1}{6}\begin{bmatrix} x & -x - 2 \\ -x^2 + 2x - 3 & 1 - x^2 \end{bmatrix}.$$

Ejemplo 8. Sea K el anillo de los enteros y

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Entonces det $A = -2$ y

$$\operatorname{adj} A = \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}.$$

Así que A no es inversible como matriz sobre el anillo de los enteros; sin embargo, se puede también considerar A como matriz sobre el anillo de los números racionales. En tal caso A es inversible y

$$A^{-1} = -\frac{1}{2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & \frac{1}{2} \end{bmatrix}.$$

En relación con matrices inversibles queremos mencionar otro hecho elemental. Las matrices semejantes tienen el mismo determinante; esto es, si P es inversible sobre K y $B = P^{-1}AP$, entonces $\det B = \det A$. Esto es claro, pues

$$\det (P^{-1}AP) = (\det P^{-1})(\det A)(\det P) = \det A.$$

Esta simple observación permite definir el determinante de un operador lineal sobre un espacio vectorial de dimensión finita. Si T es un operador lineal sobre V, se define el determinante de T como el determinante de cualquier matriz $n \times n$ que representa a T en una base ordenada de V. Como todas esas matrices son semejantes, todas tienen el mismo determinante y la definición tiene sentido. Referente a esto véase el Ejercicio 11 de la Sección 5.3.

Se va a estudiar ahora la **regla de Cramer** para resolver sistemas de ecuaciones lineales. Supóngase que A es una matriz $n \times n$ sobre el cuerpo F y que se desea resolver el sistema de ecuaciones lineales $AX = Y$ para algún n-tuple $(y_1, \ldots, y_n)$. Si $AX = Y$, entonces

$$(\operatorname{adj} A)AX = (\operatorname{adj} A)Y$$

y así

$$(\det A)X = (\operatorname{adj} A)Y.$$

Con lo que

$$\begin{aligned}(\det A)x_j &= \sum_{i=1}^{n} (\operatorname{adj} A)_{ji}y_i \\ &= \sum_{i=1}^{n} (-1)^{i+j}y_i \det A(i|j).\end{aligned}$$

Esta última expresión es el determinante de la matriz $n \times n$ que se obtiene al remplazar la columna j de A por Y. Si $\det A = 0$, nada de esto tiene sentido; sin embargo, si $\det A \neq 0$, se tiene la conocida regla de Cramer. Sea A una matriz $n \times n$ sobre el cuerpo F tal que $\det A \neq 0$. Si $y_1, \ldots, y_n$ son escalares cualesquiera de F, la solución única $X = A^{-1}Y$ del sistema de ecuaciones $AX = Y$ viene dada por

$$x_j = \frac{\det B_j}{\det A}, \qquad j = 1, \ldots, n$$

donde B_j es la matriz $n \times n$ que se obtiene de A remplazando la columna j de A por Y.

Al concluir este capítulo deseamos hacer algunos comentarios que sirvan para ubicar los determinantes en lo que se cree que es la perspectiva adecuada. De vez en cuando es necesario calcular determinantes, y esta sección se dedicó parcialmente a métodos que facilitan tal trabajo. Pero el papel principal de los determinantes en este libro es teórico. No se disputa la belleza de cuestiones como la regla de Cramer. Pero la regla de Cramer es un instrumento ineficaz para resolver sistemas de ecuaciones lineales, sobre todo porque supone demasiados cálculos. Por tanto, uno debe concentrarse en lo que dice la regla de Cramer,

más bien que en cómo calcular con ella. Ciertamente, cuando reflexionamos sobre este capítulo, esperamos que el lector insista más en entender qué es la función determinante y cómo se comporta que en cómo se calculan determinantes de matrices dadas.

Ejercicios

1. Usar la expresión de la adjunta para calcular la inversa de cada una de las siguientes matrices 3×3.

$$\begin{bmatrix} -2 & 3 & 2 \\ 6 & 0 & 3 \\ 4 & 1 & -1 \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & 0 & -\operatorname{sen}\theta \\ 0 & 1 & 0 \\ \operatorname{sen}\theta & 0 & \cos\theta \end{bmatrix}$$

2. Usar la regla de Cramer para resolver los siguientes sistemas de ecuaciones lineales sobre el cuerpo de los racionales.

(a)
$$\begin{aligned} x + y + z &= 11 \\ 2x - 6y - z &= 0 \\ 3x + 4y + 2z &= 0. \end{aligned}$$

(b)
$$\begin{aligned} 3x - 2y &= 7 \\ 3y - 2z &= 6 \\ 3z - 2x &= -1. \end{aligned}$$

3. Una matriz $n \times n$, A, sobre un cuerpo F es **antisimétrica** si $A^t = -A$. Si A es una matriz $n \times n$ antisimétrica con elementos complejos y n es impar, demostrar que $\det A = 0$.

4. Una matriz $n \times n$, A, sobre el cuerpo F se dice **ortogonal** si $AA^t = I$. Si A es ortogonal, demostrar que $\det A = \pm 1$ Dar un ejemplo de una matriz ortogonal para la cual $\det A = -1$.

5. Una matriz $n \times n$, A, sobre el cuerpo de los números complejos se dice **unitaria** si $AA^* = I$ (A^* denota la transpuesta conjugada de A). Si A es unitaria, demostrar que $|\det A| = 1$.

6. Sean T y U dos operadores lineales sobre el espacio vectorial V de dimensión finita. Demostrar que

(a) $\det(TU) = (\det T)(\det U)$.

(b) T es inversible si, y solo si, $\det T \neq 0$.

7. Sea A una matriz $n \times n$ sobre K un anillo conmutativo con unidad. Supóngase que A tiene la forma bloque:

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix}$$

donde las A_j son matrices $r_j \times r_j$. Demostrar que

$$\det A = (\det A_1)(\det A_2) \cdots (\det A_k).$$

8. Sea V el espacio vectorial de las matrices $n \times n$ sobre el cuerpo F. Sea B un elemento fijo de V y sea T_B el operador lineal sobre V definido por $T_B(A) = AB - BA$. Demostrar que $\det T_B = 0$.

9. Sea A una matriz $n \times n$ sobre un cuerpo, $A \neq 0$. Si r es un entero positivo cualquiera entre 1 y n, una **submatriz** $r \times r$ de A es cualquier matriz $r \times r$ que se obtiene suprimiendo $(n - r)$ filas y $(n - r)$ columnas de A. El **rango** de A es el mayor entero positivo r tal que alguna submatriz $r \times r$ de A tiene determinante no nulo. Demostrar que el rango de A es igual al rango de fila de A (= rango columna de A).

10. Sea A una matriz $n \times n$ sobre el cuerpo F. Demostrar que existen a lo más n escalares c distintos en F tal que $\det(cI - A) = 0$.

11. Sean A y B matrices $n \times n$ sobre el cuerpo F. Demostrar que si A es inversible existen a lo más n escalares c en F para los cuales $cA + B$ no es inversible.

12. Si V es el espacio vectorial de las matrices $n \times n$ sobre F y B es una matriz $n \times n$ dada sobre F, sean L_B y R_B los operadores lineales sobre V definidos por $L_B(A) = BA$ y $R_B(A) = AB$. Demostrar que

(a) $\det L_B = (\det B)^n$;
(b) $\det R_B = (\det B)^n$.

13. Sea V el espacio vectorial de todas las matrices $n \times n$ sobre el cuerpo de los números complejos y sea B una matriz $n \times n$ dada sobre C. Se define el operador lineal M_B en V por $M_B(A) = BAB^*$, donde $B^* = B^t$. Demostrar que

$$\det M_B = |\det B|^{2n}.$$

Sea ahora H el conjunto de todas las matrices hermíticas en V; A es hermítica si $A = A^*$. Entonces H es un espacio vectorial sobre el cuerpo de los números *reales*. Demostrar que la función T_B, definida por $T_B(A) = BAB^*$, es un operador lineal sobre el espacio vectorial real H, y luego demostrar que $T_B = |\det B|^{2n}$. (*Sugerencia*: Al calcular $\det T_B$, demostrar que V tiene una base que consta de matrices hermíticas y entonces demostrar que $\det T_B = \det M_B$.)

14. Sean A, B, C, D matrices $n \times n$, *conmutativas*, sobre el cuerpo F. Demostrar que el determinante de la matriz $2n \times 2n$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

es $\det(AD - BC)$.

5.5. *Módulos*

Si K es un anillo conmutativo con unidad, un módulo sobre K es un sistema algebraico que se comporta en forma semejante a un espacio vectorial en que K hace las veces del cuerpo escalar. Para precisar, se dice que V es un **módulo sobre** K (o un K-**módulo**) si

1. existe una adición $(\alpha, \beta) \rightarrow \alpha + \beta$ en V, respecto de la cual V es grupo conmutativo;
2. existe una multiplicación $(c, \alpha) \rightarrow c\alpha$ de elementos α en V y c en K tal que

$$\begin{aligned} (c_1 + c_2)\alpha &= c_1\alpha + c_2\alpha \\ c(\alpha_1 + \alpha_2) &= c\alpha_1 + c\alpha_2 \\ (c_1c_2)\alpha &= c_1(c_2\alpha) \\ 1\alpha &= \alpha. \end{aligned}$$

Aquí el K-módulo más importante es el de los módulos n-tuples K^n. Los módulos de matrices $K^{m\times n}$ son también importantes. Si V es cualquier módulo, se consideran combinaciones lineales, dependencia lineal e independencia lineal, tal como se hizo en un espacio vectorial. Hay que guardarse de aplicar a V cualesquiera resultados sobre un espacio vectorial que dependan de la división por escalares no nulos, operación de cuerpo que puede no ser lícita en el anillo K. Por ejemplo, si $\alpha_1, \ldots, \alpha_n$ son linealmente dependientes, no se puede concluir que algún α_i sea combinación lineal de los otros. Esto hace más difícil encontrar bases en módulos.

Una **base** para el módulo V es un subconjunto linealmente independiente que genera el módulo. Esta es la misma definición que se dio para espacios vectoriales; y la propiedad importante de una base $\mathcal{B}$ es que cada elemento de V puede expresarse unívocamente como combinación lineal de (algún número finito de) elementos de $\mathcal{B}$. Si se admite en matemática el axioma de elección (véase Apéndice), se puede demostrar que todo espacio vectorial tiene una base. El lector está bien enterado de que existe una base en todo espacio vectorial generado por un número finito de vectores. Pero este no es el caso para los módulos. Por ello se necesitan nombres especiales para los módulos que tienen base y los que son generados por un número finito de elementos.

Definición. *El K-módulo se dice* **módulo libre** *si tiene una base. Si V tiene una base finita de n elementos, entonces V se dice un* K-**módulo libre con n generadores.**

Definición. *El módulo V es* **finitamente generado** *si tiene un subconjunto finito que genere V. El* **rango** *de un módulo finitamente generado es el menor entero k tal que los k elementos generen V.*

Se repite que un módulo puede ser finitamente generado sin tener una base finita. Si V es un K-módulo libre con n-generadores, entonces V es isomorfo al módulo K^n. Si $\{\beta_1, \ldots, \beta_n\}$ es una base de V, existe un isomorfismo que aplica el vector $c_1\beta_1 + \cdots + c_n\beta_n$ sobre el n-tuple $(c_1, \ldots, c_n)$ de K^n. No es evidente, en forma inmediata, que el mismo módulo V no pueda ser también un módulo libre con k generadores, con $k \neq n$. Es decir, no es obvio que dos bases cualesquiera de V deban tener el mismo número de elementos. La demostración de esto es una interesante aplicación de los determinantes.

Teorema 5. *Sea K un anillo conmutativo con unidad. Si V es un K-módulo libre con n generadores, entonces el rango de V es n.*

Demostración. Se tiene que demostrar que V no puede ser generado por menos de n de sus elementos. Como V es isomorfo a K^n, se debe hacer ver que, si $m < n$, el módulo K^n no es generado por los n-tuples $\alpha_1, \ldots, \alpha_m$. Sea A la matriz de filas $\alpha_1, \ldots, \alpha_m$. Supóngase que cada uno de los vectores de la base canónica $\epsilon_1, \ldots, \epsilon_n$ sea combinación lineal de los $\alpha_1, \ldots, \alpha_m$. Entonces existe una matriz P en $K^{n\times m}$ tal que

$$PA = I$$

donde I es la matriz identidad $n \times n$. Sea $\tilde{A}$ la matriz $n \times n$ que se obtiene al adjuntar $n - m$ filas de 0 a la parte inferior de A y sea $\tilde{P}$ cualquier matriz $n \times n$ que tiene las columnas de P como sus primeras n columnas. Entonces

$$\tilde{P}\tilde{A} = I.$$

Por consiguiente, $\det \tilde{A} \neq 0$. Pero como $m < n$, al menos una fila de $\tilde{A}$ tiene todos los elementos 0. Esta contradicción muestra que $\alpha_1, \ldots, \alpha_m$ no generan K^n. ▌

Es interesante anotar que el Teorema 5 establece la unicidad de la dimensión de un espacio vectorial (de dimensión finita). La demostración, basada en la existencia de la función determinante, es bastante diferente de la demostración dada en el Capítulo 2. Por el Teorema 5, se sabe que «módulo libre de rango n» es lo mismo que el «módulo libre con n generadores».

Si V es un módulo libre sobre K, el **módulo dual** V^* consta de todas las funciones lineales f de V en K. Si V es un módulo libre de rango n, entonces V^* es también un módulo libre de rango n. La demostración es la misma que para espacios vectoriales. Si $\{\beta_1, \ldots, \beta_n\}$ es una base ordenada de V, existe una **base dual** $\{f_1, \ldots, f_n\}$ asociada del módulo V^*. La función f_i asigna a cada α de V su i-ésima coordenada respecto de $\{\beta_1, \ldots, \beta_n\}$

$$\alpha = f_1(\alpha)\beta_1 + \cdots + f_n(\alpha)\beta_n.$$

Si f es una función lineal sobre V, entonces

$$f = f(\beta_1)f_1 + \cdots + f(\beta_n)f_n.$$

5.6. *Funciones multilineales*

El propósito de esta sección es colocar el estudio de los determinantes en lo que se cree que es la perspectiva adecuada. Se estudiarán las formas multilineales alternantes sobre módulos. Estas formas son la generalización natural de los determinantes según los presentamos. El lector que no haya leído (o no desee leer) la breve introducción sobre módulos de la Sección 5.5, puede aún estudiar con provecho esta sección, leyendo cada vez «espacio vectorial sobre F de dimensión n» en vez de «módulo libre sobre K de rango n'».

Sea K un anillo conmutativo con unidad y sea V un módulo sobre K. Si r es un entero positivo, una función L de $V^r = V \times V \times \cdots \times V$ en K se dice **multilineal** si $L(\alpha_1, \ldots, \alpha_r)$ es lineal como función de cada α_i, cuando los otros α_j se dejan fijos; esto es, si para cada i

$$L(\alpha_1, \ldots, c\alpha_i + \beta_i, \ldots, \alpha_r) = cL(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_r + L(\alpha_1, \ldots, \beta_i, \ldots, \alpha_r).$$

Una función multilineal en V^r se llama también **forma r-lineal** sobre V o **forma multilineal de grado r** sobre V. Tales funciones se llaman **r-tensores** sobre V.

La colección de todas las funciones multilineales sobre V^r se denotará $M^r(V)$. Si L y M pertenecen a $M^r(V)$, entonces la suma $L + M$:

$$(L + M)(\alpha_1, \ldots, \alpha_r) = L(\alpha_1, \ldots, \alpha_r) + M(\alpha_1, \ldots, \alpha_r)$$

es también multilineal; y, si c es un elemento de K, el producto cL:

$$(cL)(\alpha_1, \ldots, \alpha_r) = cL(\alpha_1, \ldots, \alpha_r)$$

es multilineal. Por tanto, $M^r(V)$ es un K-módulo —un submódulo del módulo de todas las funciones de V^r en K.

Si $r = 1$, se tiene $M'(V) = V^*$, ei módulo dual de funciones lineales sobre V. Las funciones lineales pueden ser también usadas para construir ejemplos de formas multilineales de órdenes más elevados. Si $f_1, \ldots, f_r$ son funciones lineales sobre V, defínase

$$L(\alpha_1, \ldots, \alpha_r) = f_1(\alpha_1)f_2(\alpha_2) \cdots f_r(\alpha_r).$$

Evidentemente, L es una forma r-lineal en V.

Ejemplo 9. Si V es un módulo, una forma 2-lineal sobre V se llama a menudo **forma bilineal** sobre V. Sea A una matriz $n \times n$ con elementos en K. Entonces

$$L(X, Y) = Y^tAX$$

define una forma bilineal L sobre el módulo $K^{n \times 1}$. Análogamente,

$$M(\alpha, \beta) = \alpha A \beta^t$$

define una forma bilineal M sobre K^n.

Ejemplo 10. La función determinante asocia a cada matriz $n \times n$, A, un elemento det A en K. Si det A es considerado como función de las filas de A:

$$\det A = D(\alpha_1, \ldots, \alpha_n)$$

entonces D es una forma n-lineal en K^n.

Ejemplo 11. Es fácil lograr una expresión algebraica de la forma r-lineal general sobre el módulo K^n. Si $\alpha_1, \ldots, \alpha_r$ son vectores de V y A es la matriz $r \times n$ con filas $\alpha_1, \ldots, \alpha_r$, entonces, para cualquier función L de $M^r(K^n)$,

$$\begin{aligned}
L(\alpha_1, \ldots, \alpha_r) &= L\left(\sum_{j=1}^{n} A_{1j}\epsilon_j, \alpha_2, \ldots, \alpha_r\right) \\
&= \sum_{j=1}^{n} A_{1j}L(\epsilon_j, \alpha_2, \ldots, \alpha_r) \\
&= \sum_{j=1}^{n} A_{1j}L\left(\epsilon_j, \sum_{j=1}^{n} A_{2k}\epsilon_k, \ldots, \alpha_r\right) \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n} A_{1j}A_{2k}L(\epsilon_j, \epsilon_k, \alpha_3, \ldots, \alpha_r) \\
&= \sum_{j,k=1}^{n} A_{1j}A_{2k}L(\epsilon_j, \epsilon_k, \alpha_3, \ldots, \alpha_r).
\end{aligned}$$

Si se remplazan $\alpha_1, \ldots, \alpha_r$ sucesivamente por sus expresiones como combinaciones lineales de los vectores de la base canónica y si se escribe $A(i, j)$ por A_{ij}, se obtiene:

$$(5\text{-}26) \qquad L(\alpha_1, \ldots, \alpha_r) = \sum_{j_1, \ldots, j_r = 1}^{n} A(1, j_1) \cdots A(r, j_r) L(\epsilon_{j_1}, \ldots \epsilon_{j_r}).$$

En (5-26) hay un término por cada n-tuple $J = (j_1, \ldots, j_r)$ de enteros positivos entre 1 y n. Hay n^r de tales r-tuples. Así, pues, L está completamente determinada por (5-26) y los valores particulares

$$c_J = L(\epsilon_{j_1}, \ldots, \epsilon_{j_r})$$

asignados a los n^r elementos $(\epsilon_{j_1}, \ldots, \epsilon_{j_r})$. Es también fácil ver que si para cada r-tuple J se elige un elemento c_j de K, entonces

$$(5\text{-}27) \qquad L(\alpha_1, \ldots, \alpha_r) = \sum_J A(1, j_1) \cdots A(r, j_r) c_J$$

define una forma r-lineal sobre K^n.

Supóngase que L es una función multilineal sobre V^r y M es una función multilineal sobre V^s. Se define una función $L \otimes M$ sobre V^{r+s} por

$$(5\text{-}28) \qquad (L \otimes M)(\alpha_1, \ldots, \alpha_{r+s}) = L(\alpha_1, \ldots, \alpha_r) M(\alpha_{r+1}, \ldots, \alpha_{r+s}).$$

Si se piensa de V^{r+s} como $V^r \times V^s$, entonces para α en V^r y β en V^s

$$(L \otimes M)(\alpha, \beta) = L(\alpha) M(\beta).$$

Es claro que $L \otimes M$ es multilineal sobre V^{r+s}. La función $L \otimes M$ se llama **producto tensorial** de L y M. El producto tensorial no es conmutativo. En efecto, $M \otimes L \neq L \otimes M$, a menos que $L = 0$ o $M = 0$; sin embargo, el producto tensorial se relaciona perfectamente con las operaciones modulares en M^r y M^s.

Lema. *Sean L, L_1 formas r-lineales sobre V; sean M, M_1 formas s-lineales sobre V, y sea c un elemento de K.*

(a) $(cL + L_1) \otimes M = c(L \otimes M) + L_1 \otimes M$;
(b) $L \otimes (cM + M_1) = c(L \otimes M) + L \otimes M_1$.

Demostración. Se deja como ejercicio.

El producto tensorial es asociativo, es decir, si L, M y N son (respectivamente) formas r-, s- y t-lineales sobre V, entonces

$$(L \otimes M) \otimes N = L \otimes (M \otimes N).$$

Ello es inmediato por ser la multiplicación en K asociativa. Por tanto, si L_1, $L_2, \ldots, L_k$ son funciones multilineales sobre $V^{r_1}, \ldots, V^{r_k}$, entonces el producto tensorial

$$L = L_1 \otimes \cdots \otimes L_k$$

está definido, en forma inequívoca, como función multilineal sobre V^r, donde $r = r_1 + \cdots + r_k$. Se mencionó ya un caso particular al respecto. Si $f_1, \ldots, f_r$ son funciones lineales sobre V, entonces el producto tensorial

$$L = f_1 \otimes \cdots \otimes f_r$$

viene dado por

$$L(\alpha_1, \ldots, \alpha_r) = f_1(\alpha_1) \cdots f_r(\alpha_r).$$

Teorema 6. *Sea K un anillo conmutativo con unidad. Si V es un K-módulo libre de rango n, entonces $M^r(V)$ es un K-módulo libre de rango n^r; en efecto, si $\{f_1, \ldots, f_n\}$ es una base para el módulo dual V^*, los n^r productos tensoriales*

$$f_{j_1} \otimes \cdots \otimes f_{j_r}, \qquad 1 \leq j_1 \leq n, \ \ldots, \ 1 \leq j_r \leq n$$

forman una base de $M^r(V)$.

Demostración. Sea $\{f_1, \ldots, f_n\}$ una base ordenada de V^* que es dual de la base $\{\beta_1, \ldots, \beta_n\}$ de V. Para todo vector α de V se tiene

$$\alpha = f_1(\alpha)\beta_1 + \cdots + f_n(\alpha)\beta_n.$$

Se hacen ahora los cálculos realizados en el Ejemplo 11. Si L es una r-forma lineal sobre V y $\alpha_1, \ldots, \alpha_r$ son elementos de V, entonces por (5-26)

$$L(\alpha_1, \ldots, \alpha_r) = \sum_{j_1, \ldots, j_r} f_{j_1}(\alpha_1) \cdots f_{j_r}(\alpha_r) L(\beta_{j_1}, \ldots, \beta_{j_r}).$$

En otras palabras,

$$L = \sum_{j_1, \ldots, j_r} L(\beta_{j_1}, \ldots, \beta_{j_r}) f_{j_1} \otimes \cdots \otimes f_{j_r}. \tag{5-29}$$

Esto muestra que los n^r productos tensoriales

$$E_J = f_{j_1} \otimes \cdots \otimes f_{j_r} \tag{5-30}$$

dados por los r-tuples $J = (j_1, \ldots, j_r)$ generan el módulo $M^r(V)$. Se ve que las diversas r-formas E_J son independientes como sigue. Supóngase que para cada J tenemos un elemento c_J en K y formamos la función multilineal

$$L = \sum_J c_J E_J. \tag{5-31}$$

Obsérvese que si $I = (i_1, \ldots, i_r)$, entonces

$$E_J(\beta_{i_1}, \ldots, \beta_{i_r}) = \begin{cases} 0, & I \neq J \\ 1, & I = J. \end{cases}$$

Por tanto, se ve por (5-31) que

$$c_I = L(\beta_{i_1}, \ldots, \beta_{i_r}). \tag{5-32}$$

En particular, si $L = 0$, entonces $c_I = 0$ para todo r-tuple I. ∎

Definición. *Sea L una forma r-lineal sobre un K-módulo V. Se dice que L es* **alternada** *si $L(\alpha_1, \ldots, \alpha_r) = 0$, siempre que $\alpha_i = \alpha_j$, con $i \neq j$.*

Si L es una función multilineal alternada sobre V^r, entonces

$$L(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_j, \ldots, \alpha_r) = -L(\alpha_1, \ldots, \alpha_j, \ldots, \alpha_i, \ldots, \alpha_r).$$

Es decir, si se transponen dos de los vectores (con índices diferentes) del r-tuple $(\alpha_1, \ldots, \alpha_r)$, el valor asociado de L cambia de signo. Como toda permutación σ es un producto de transposiciones, se ve que $L(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma r}) = (\operatorname{sgn} \sigma)\, L(\alpha_1, \ldots, \alpha_r)$.

Se designa por $\Lambda^r(V)$ la colección de todas las formas r-lineales sobre V. Es fácil ver que $\Lambda^r(V)$ es un submódulo de $M^r(V)$.

Ejemplo 12. Al comienzo de este capítulo se vio que sobre el módulo K^n existe precisamente una forma alternada n-lineal D con la propiedad de que $D(\epsilon_1, \ldots, \epsilon_n) = 1$. También se demostró, en el Teorema 2, que si L es cualquier forma de $\Lambda^n(K^n)$, entonces

$$L = L(\epsilon_1, \ldots, \epsilon_n)D.$$

O sea que $\Lambda^n(K^n)$ es un K-módulo libre de rango 1. También se desarrolló una fórmula explícita (5-15) para D. En términos de la notación que ahora estamos usando, dicha fórmula se puede escribir

$$D = \sum_\sigma (\operatorname{sgn} \sigma) f_{\sigma 1} \otimes \cdots \otimes f_{\sigma n}. \tag{5-33}$$

donde $f_1, \ldots, f_n$ son las funciones coordenadas canónicas sobre K^n y la suma se extiende sobre las $n!$ permutaciones diferentes σ del conjunto $\{1, \ldots, n\}$. Si escribimos el determinante de una matriz A como

$$\det A = \sum_\sigma (\operatorname{sgn} \sigma)\, A(\sigma 1, 1) \cdots A(\sigma n, n)$$

entonces se obtiene una expresión diferente de D:

$$\begin{aligned} D(\alpha_1, \ldots, \alpha_n) &= \sum_\sigma (\operatorname{sgn} \sigma) f_1(\alpha_{\sigma 1}) \cdots f_n(\alpha_{\sigma n}) \\ &= \sum_\sigma (\operatorname{sgn} \sigma)\, L(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma n}) \end{aligned} \tag{5-34}$$

donde $L = f_1 \otimes \cdots \otimes f_n$.

Hay un método general para asociar una forma alternada a una forma multilineal. Si L es una forma r-lineal sobre un módulo V y si σ es una permutación de $\{1, \ldots, r\}$, se tiene otra función r-lineal L_σ definiendo

$$L_\sigma(\alpha_1, \ldots, \alpha_r) = L(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma r}).$$

Si L resulta alternada, entonces $L_\sigma = (\operatorname{sgn} \sigma)L$. Ahora, para cada L en $M^r(V)$ se define una función $\pi_r L$ en $M^r(V)$ por

$$\pi_r L = \sum_\sigma (\operatorname{sgn} \sigma) L_\sigma \tag{5-35}$$

esto es,

(5-36) $$(\pi_r L)(\alpha_1, \ldots, \alpha_r) = \sum_\sigma (\text{sgn}\,\sigma)\, L(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma r}).$$

Lema. *π_r es una transformación lineal de $M^r(V)$ en $\Lambda^r(V)$. Si L está en $\Lambda^r(V)$, entonces $\pi_r L = r!L$.*

Demostración. Sea τ una permutación cualquiera de $\{1, \ldots, r\}$. Entonces

$$\begin{aligned}(\pi_r L)(\alpha_{\tau 1}, \ldots, \alpha_{\tau r}) &= \sum_\sigma (\text{sgn}\,\sigma)\, L(\alpha_{\tau\sigma 1}, \ldots, \alpha_{\tau\sigma r})\\ &= (\text{sgn}\,\tau) \sum_\sigma (\text{sgn}\,\tau\sigma)\, L(\alpha_{\tau\sigma 1}, \ldots, \alpha_{\tau\sigma r}).\end{aligned}$$

Como σ recorre (una vez) todas las permutaciones de $\{1, \ldots, r\}$, también lo hace $\tau\sigma$. Por tanto,

$$(\pi_r L)(\alpha_{\tau 1}, \ldots, \alpha_{\tau r}) = (\text{sgn}\,\tau)(\pi_r L)(\alpha_1, \ldots, \alpha_r).$$

Así, $\pi_r L$ es una forma alternada.

Si L está en $\Lambda^r(V)$, entonces $L(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma r}) = (\text{sgn}\,\sigma)\, L(\alpha_1, \ldots, \alpha_r)$ para todo σ; luego $\pi_r L = r!L$. ▌

En (5-33) se demostró que la función determinante D en $\Lambda^n(K^n)$ es

$$D = \pi_n(f_1 \otimes \cdots \otimes f_n)$$

donde $f_1, \ldots, f_n$ son las funciones coordenadas canónicas sobre K^n. Hay una observación importante que hacer en relación con este último lema. Si K es un cuerpo de característica cero, tal que $r!$ es inversible en K, entonces π aplica $M^r(V)$ sobre $\Lambda^r(V)$. En realidad, en ese caso es más natural desde cierto punto de vista usar la aplicación $\pi_1 = (1/r!)\pi$ en vez de π, ya que π_1 es una proyección de $M^r(V)$ sobre $\Lambda^r(V)$; es decir, una aplicación lineal de $M^r(V)$ sobre $\Lambda^r(V)$ tal que $\pi_1(L) = L$ si, y solo si, L está en $\Lambda^r(V)$.

Teorema 7. *Sea K un anillo conmutativo con unidad y sea V un K-módulo libre de rango n. Si $r > n$, entonces $\Lambda^r(V) = \{0\}$. Si $1 \leq t \leq n$, entonces $\Lambda^r(V)$ es un K-módulo libre de rango $\binom{n}{r}$.*

Demostración. Sea $\{\beta_1, \ldots, \beta_n\}$ una base ordenada de V con base dual $\{f_1, \ldots, f_n\}$. Si L pertenece a $M^r(V)$, se tiene

(5-37) $$L = \sum_J L(\beta_{j_1}, \ldots, \beta_{j_r}) f_{j_1} \otimes \cdots \otimes f_{j_r}$$

donde la suma se extiende sobre todos los r-tuples $J = (j_1, \ldots, j_n)$ de enteros comprendidos entre 1 y n. Si L es alternada, entonces

$$L(\beta_{j_1}, \ldots, \beta_{j_r}) = 0$$

siempre que dos de los índices j_i sean iguales. Si $r > n$, entonces en cada r-tuple J algún entero debe estar repetido. Con lo cual $\Lambda^r(V) = \{0\}$ si $r > n$.

Supóngase ahora que $1 \leq r \leq n$. Si L está en $\Lambda^r(V)$, la suma en (5-37) debe extenderse solo sobre los r-tuples J para los que $j_1, \ldots, j_r$ son distintos, ya que todos los otros términos son 0. Cada r-tuple de enteros distintos entre 1 y n es una permutación de un r-tuple $J = (j_1, \ldots, j_r)$ tal que $j_1 < \cdots < j_r$. Este tipo especial de r-tuple se llamará una combinación **r-aria** de $\{1, \ldots, n\}$. Hay

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

de tales combinaciones.

Supóngase que se fije una combinación r-aria J. Sea L_J la suma de todos los términos en (5-37) que corresponden a las permutaciones de las combinaciones J. Si σ es una permutación de $\{1, \ldots, r\}$, entonces

$$L(\beta_{j_{\sigma 1}}, \ldots, \beta_{j_{\sigma r}}) = (\operatorname{sgn} \sigma)\, L(\beta_{j_1}, \ldots, \beta_{j_r}).$$

Con lo que

(5-38) $$L_J = L(\beta_{j_1}, \ldots, \beta_{j_r}) D_J$$

donde

(5-39) $$\begin{aligned} D_J &= \sum_{\sigma} (\operatorname{sgn} \sigma)\, f_{j_{\sigma 1}} \otimes \cdots \otimes f_{j_{\sigma r}} \\ &= \pi_r(f_{j_1} \otimes \cdots \otimes f_{j_r}). \end{aligned}$$

Se ve de (5-39) que cada D_J es alternada y que

(5-40) $$L = \sum_{\text{combinación } J} L(\beta_{j_1}, \ldots, \beta_{j_r}) D_J$$

para todo L de $\Lambda^r(V)$. La aserción es que las $\binom{n}{r}$ formas D_J constituyen una base de $\Lambda^r(V)$. Hemos visto que generan $\Lambda^r(V)$ y es fácil ver que son independientes. Si $I = (i_1, \ldots, i_r)$ y $J = (j_1, \ldots, j_r)$ son combinaciones, entonces

(5-41) $$D_J(\beta_{i_1}, \ldots, \beta_{i_r}) = \begin{cases} 1, & I = J \\ 0, & I \neq J \end{cases}.$$

Supóngase que se tiene para cada combinación un escalar c_J y defínase

$$L = \sum_J c_J D_J.$$

De (5-40) y (5-41) se tiene

$$c_I = L(\beta_{i_1}, \ldots, \beta_{i_r}).$$

En particular, si $L = 0$, entonces $c_I = 0$ para cada combinación I. ∎

Corolario. *Si V es un K-módulo libre de rango n, entonces $\Lambda^n(V)$ es un K-módulo libre de rango 1. Si T es un operador lineal sobre V, existe un único elemento c en K tal que*

$$L(T\alpha_1, \ldots, T\alpha_n) = cL(\alpha_1, \ldots, \alpha_n)$$

para cada forma n-lineal alternada de L sobre V.

Demostración. Si L está en $\Lambda^n(V)$, entonces evidentemente

$$L_T(\alpha_1, \ldots, \alpha_n) = L(T\alpha_1, \ldots, T\alpha_n)$$

define una forma n-lineal alternada L_T. Sea M un generador para el rango 1 módulo $\Lambda^n(V)$. Todo L en $\Lambda^n(V)$ es unívocamente expresable como $L = aM$ para algún a en K. En particular, $M_T = cM$ para un cierto c. Para $L = aM$ se tiene

$$\begin{aligned} L_T &= (aM)_T \\ &= aM_T \\ &= a(cM) \\ &= c(aM) \\ &= cL. \quad \blacksquare \end{aligned}$$

Naturalmente, el elemento c del último corolario es llamado el **determinante** de T. De (5-39), para el caso $r = n$ (cuando hay solo una combinación $J = (1, \ldots, n)$), se ve que el determinante de T es el determinante de la matriz que representa T en cualquier base ordenada $\{\beta_1, \ldots, \beta_n\}$. Veamos por qué. La matriz representante tiene i, j elementos

$$A_{ij} = f_j(T\beta_i)$$

de modo que

$$\begin{aligned} D_J(T\beta_1, \ldots, T\beta_n) &= \sum_\sigma (\operatorname{sgn} \sigma) A(1, \sigma 1) \cdots A(n, \sigma n) \\ &= \det A. \end{aligned}$$

Por otro lado,

$$\begin{aligned} D_J(T\beta_1, \ldots, T\beta_n) &= (\det T)\, D_J(\beta_1, \ldots, \beta_n) \\ &= \det T. \end{aligned}$$

La razón de estas observaciones es que, por medio del Teorema 7 y su corolario, se obtiene una definición del determinante de un operador lineal que no supone el conocimiento de los determinantes de matrices. Los determinantes de matrices pueden ser definidos en términos de determinantes de operadores, en vez de al contrario.

Queremos decir algo más con respecto a las formas especiales r-lineales alternadas D_J que se asociaron a una base $\{f_1, \ldots, f_n\}$ de V^* en (5-39). Es importante entender que $D_J(\alpha_1, \ldots, \alpha_r)$ es el determinante de cierta matriz $r \times r$. Si

$$A_{ij} = f_j(\alpha_i), \qquad 1 \leq i \leq r, 1 \leq j \leq n,$$

esto es, si

$$\alpha_i = A_{i1}\beta_1 + \cdots + A_{in}\beta_n, \qquad 1 \leq i \leq r$$

y J es la combinación r-aria $(j_1, \ldots, j_r)$, entonces

(5-42)
$$\begin{aligned} D_J(\alpha_1, \ldots, \alpha_r) &= \sum_\sigma (\operatorname{sgn} \sigma) A(1, j_{\sigma 1}) \cdots A(n, j_{\sigma n}) \\ &= \det \begin{bmatrix} A(1, j_1) & \cdots & A(1, j_r) \\ \vdots & & \vdots \\ A(r, j_1) & \cdots & A(r, j_r) \end{bmatrix}. \end{aligned}$$

Así que $D_J(\alpha_1, \ldots, \alpha_r)$ es el determinante de la matriz $r \times r$ formada por las columnas $j_1, \ldots, j_r$ de la matriz $r \times n$ que tiene (los n-tuples coordenados de) $\alpha_1, \ldots, \alpha_r$ como sus filas. Otra notación que se usa a veces para este determinante es

$$(5\text{-}43) \qquad D_J(\alpha_1, \ldots, \alpha_r) = \frac{\partial(\alpha_1, \ldots, \alpha_r)}{\partial(\beta_{j_1}, \ldots, \beta_{j_r})}.$$

En esta notación la demostración del Teorema 7 muestra que toda forma r-lineal alternada L puede expresarse respecto a una base $\{\beta_1, \ldots, \beta_n\}$ por la igualdad

$$(5\text{-}44) \qquad L(\alpha_1, \ldots, \alpha_r) = \sum_{j_1 < \cdots < j_r} \frac{\partial(\alpha_1, \ldots, \alpha_r)}{\partial(\beta_{j_1}, \ldots, \beta_{j_r})} L(\beta_{j_1}, \ldots, \beta_{j_r}).$$

5.7. *El anillo de Grassman*

Muchas de las propiedades importantes de los determinantes y de las formas multilineales alternadas se describen mejor mediante una operación de multiplicación sobre formas llamada producto exterior. Si L y M son, respectivamente, formas r- y s-lineales alternadas sobre el módulo V, se tiene un producto asociado a L y K, el producto tensorial $L \otimes M$. Esta no es una forma alternada, a menos que $L = 0$ o $M = 0$; sin embargo, se tiene una manera natural de proyectarlo en $\Lambda^{r+s}(V)$. Parecería que

$$(5\text{-}45) \qquad L \cdot M = \pi_{r+s}(L \otimes M)$$

fuera la multiplicación «natural» de las formas alternadas. Pero ¿lo es?

Consideremos un ejemplo concreto. Supóngase que V es el módulo K^n y que $f_1, \ldots, f_n$ sean las funciones coordenadas canónicas sobre K^n. Si $i \neq j$, entonces

$$f_i \cdot f_j = \pi_2(f_i \otimes f_j)$$

es la función (determinante)

$$D_{ij} = f_i \otimes f_j - f_j \otimes f_i$$

dada por (5-39). Supóngase ahora que k es un índice diferente de i y de j. Entonces

$$\begin{aligned} D_{ij} \cdot f_k &= \pi_3[(f_i \otimes f_j - f_j \otimes f_i) \otimes f_k] \\ &= \pi_3(f_i \otimes f_j \otimes f_k) - \pi_3(f_j \otimes f_i \otimes f_k). \end{aligned}$$

La demostración del lema que sigue a la ecuación (5-36) muestra que para cualquier forma r-lineal L y para cualquier permutación σ de $\{1, \ldots, r\}$

$$\pi_r(L_\sigma) = \operatorname{sgn} \sigma\, \pi_r(L)$$

Luego, $D_{ij} \cdot f_k = 2\pi_3(f_i \otimes f_j \otimes f_k)$. Por un cálculo semejante $f_i \cdot D_{jk} = 2\pi_3(f_i \otimes f_j \otimes f_k)$. Con lo que

$$(f_i \cdot f_j) \cdot f_k = f_i \cdot (f_j \cdot f_k)$$

todo lo cual parece muy prometedor. Pero hay un inconveniente. A pesar del cálculo que se ha hecho, la supuesta multiplicación de (5-45) no es asociativa. En efecto, si l es un índice distinto de los i, j, k, entonces se puede calcular que

$$D_{ij} \cdot D_{kl} = 4\pi_4(f_i \otimes f_j \otimes f_k \otimes f_l)$$

y que

$$(D_{ij} \cdot f_k) \cdot f_l = 6\pi_4(f_i \otimes f_j \otimes f_k \otimes f_l).$$

Con lo que, en general

$$(f_i \cdot f_j) \cdot (f_k \cdot f_l) \neq [(f_i \cdot f_j) \cdot f_k] \cdot f_l$$

y se ve que el primer intento de encontrar una multiplicación ha dado una operación no asociativa.

El lector no debe sorprenderse si encuentra más bien fatigoso hacer una verificación directa para ver que las dos ecuaciones no son asociativas. Ello es típico de la materia, y también es típico que hay un hecho general que simplifica considerablemente las operaciones.

Supóngase que L es una forma r-lineal y que M es una forma s-lineal sobre el módulo V. Entonces

$$\begin{aligned}\pi_{r+s}((\pi_r L) \otimes (\pi_s M)) &= \pi_{r+s}(\sum_{\sigma,\tau} (\text{sgn } \sigma)(\text{sgn } \tau)L_\sigma \otimes M_\tau) \\ &= \sum_{\sigma,\tau} (\text{sgn } \sigma)(\text{sgn } \tau)\pi_{r+s}(L_\sigma \otimes M_\tau)\end{aligned}$$

donde σ varía sobre el grupo simétrico, S_r, de todas las permutaciones de $\{1, \ldots, r\}$, y τ varía sobre S_s. Cada par σ, τ define un elemento (σ, τ) de S_{r+s} que permuta los primeros r elementos de $\{1, \ldots, r + s\}$ según σ, y los últimos s elementos según τ. Es claro que

$$\text{sgn } (\sigma, \tau) = (\text{sgn } \sigma)(\text{sgn } \tau)$$

y que

$$(L \otimes M)_{(\sigma,\tau)} = L_\sigma \otimes L_\tau.$$

Por tanto,

$$\pi_{r+s}[(\pi_r L) \otimes (\pi_s M)] = \sum_{\sigma,\tau} \text{sgn } (\sigma, \tau)\ \pi_{r+s}\ [(L \otimes M)_{(\sigma,\tau)}].$$

Ahora bien, ya se ha observado que

$$\text{sgn } (\sigma, \tau)\pi_{r+s}[(L \otimes M)_{(\sigma,\tau)}] = \pi_{r+s}(L \otimes M).$$

Así, pues, se sigue que

$$(5\text{-}46) \qquad \pi_{r+s}[(\pi_r L) \otimes (\pi_s M)] = r!s!\ \pi_{r+s}(L \otimes M).$$

Esta fórmula simplifica numerosos cálculos. Por ejemplo, supóngase que se tiene una combinación r-aria $I = (i_1, \ldots, i_r)$ y una combinación s-aria $J = (j_1, \ldots, j_s)$. Para simplificar, supóngase, además, que

$$i_1 < \cdots < i_r < j_1 < \cdots < j_s.$$

Entonces se tienen las funciones determinantes asociadas

$$D_I = \pi_r(E_I)$$
$$D_J = \pi_s(E_J)$$

donde E_I y E_J están dados por (5-30). Usando (5-46), se ve inmediatamente que

$$\begin{aligned} D_I \cdot D_J &= \pi_{r+s}[\pi_r(E_I) \otimes \pi_s(E_J)] \\ &= r!s!\pi_{r+s}(E_I \otimes E_J). \end{aligned}$$

Como $E_I \otimes E_J = E_{I \cup J}$, se sigue que

$$D_I \cdot D_J = r!s!\, D_{I \cup J}.$$

Esto sugiere que la ausencia de asociatividad en la multiplicación (5-45) resulta del hecho de que $D_I \cdot D_J \neq D_{I \cup J}$. Después de todo, el producto de D_I y D_J debería ser $D_{I \cup J}$. Para remediar esta situación, definiremos un nuevo producto, el **producto exterior** de una forma r-lineal alternada L y una forma s-lineal alternada M por

$$L \wedge M = \frac{1}{r!s!} \pi_{r+s}(L \otimes M). \tag{5-47}$$

Se tiene entonces

$$D_I \wedge D_J = D_{I \cup J}$$

para las funciones determinantes sobre K^n, y, si es que hay justicia después de todo, se ha tenido que lograr la multiplicación apropiada de las formas multilineales alternadas. Desafortunadamente, (5-47) deja de tener sentido para el caso más general en consideración, ya que no es posible dividir por $r!s!$ en el anillo K. Si K es un cuerpo de característica cero, entonces (5-47) tiene sentido y se puede proceder sin más a demostrar que el producto exterior es asociativo.

Teorema 8. *Sea K un cuerpo de característica cero, y V un espacio vectorial sobre K. Entonces el producto exterior es una operación asociativa sobre las formas multilineales alternadas sobre V. En otras palabras, si L, M y N son formas multilineales alternadas sobre V de grados r, s y t, respectivamente, entonces*

$$(L \wedge M) \wedge N = L \wedge (M \wedge N).$$

Demostración. De (5-47) se sigue que $cd(L \wedge M) = cL \wedge dM$ para cualesquiera escalares c y d. Luego

$$r!s!t![(L \wedge M) \wedge N] = r!s!(L \wedge M) \wedge t!N$$

y como $\pi_t(N) - t!N$, se tiene que

$$\begin{aligned} r!s!t![(L \wedge M) \wedge N] &= \pi_{r+s}(L \otimes M) \wedge \pi_t(N) \\ &= \frac{1}{(r+s)!}\frac{1}{t!} \pi_{r+s+t}[\pi_{r+s}(L \otimes M) \otimes \pi_t(N)]. \end{aligned}$$

Por (5-46) se sabe que

$$r!s!t[(L \wedge M) \wedge N] = \pi_{r+s+t}(L \otimes M \otimes N).$$

Por un cálculo análogo

$$r!s!t![L \wedge (M \wedge N)] = \pi_{r+s+t}(L \otimes M \otimes N)$$

y, por tanto, $(L \wedge M) \wedge N = L \wedge (M \wedge N)$. ▌

Volvamos ahora al caso general en el que solo se supone que K es un anillo conmutativo con unidad. El primer problema es remplazar (5-47) por una definición equivalente que opere en general. Si L y M son formas multilineales alternadas de grado r y s, respectivamente, se construirá una forma canónica multilineal alternada $L \wedge M$ de grado $r + s$ tal que

$$r!s!(L \wedge M) = \pi_{r+s}(L \otimes M).$$

Recordemos cómo se define $\pi_{r+s}(L \otimes M)$. A cada permutación σ de $\{1, \ldots, r + s\}$ se asocia la función multilineal

(5-48) $$(\operatorname{sgn} \sigma)(L \otimes M)_\sigma$$

donde

$$(L \otimes M)_\sigma(\alpha_1, \ldots, \alpha_{r+s}) = (L \otimes M)(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma(r+s)})$$

y se suman las funciones (5-48) sobre todas las permutaciones σ. Hay $(r + s)!$ permutaciones; pero como L y M son alternadas, muchas de las funciones (5-48) son una misma. En realidad, hay a lo más

$$\frac{(r+s)!}{r!s!}$$

funciones (5-48) distintas. Veamos por qué. Sea S_{r+s} el conjunto de las permutaciones de $\{1, \ldots, r + s\}$, es decir, sea S_{r+s} el grupo simétrico de grado $r + s$. Como en la demostración de (5-46), se distingue el subconjunto G que consta de las permutaciones σ que permutan los conjuntos $\{1, \ldots, r\}$ y $\{r + 1, \ldots, r + s\}$ sobre sí mismos. En otras palabras, σ pertenece a G si $1 \leq \sigma i \leq r$ para todo i entre 1 y r. (Necesariamente se sigue que $r + 1 \leq \sigma i \leq r + s$ para todo j entre $r + 1$ y $r + s$.) Ahora G es un subgrupo de S_{r+s}, esto es, si σ y τ están en G, entonces $\sigma\tau^{-1}$ está en G. Evidentemente, G tiene $r!s!$ elementos.

Se tiene una aplicación

$$S_{r+s} \xrightarrow{\psi} M^{r+s}(V)$$

definida por

$$\psi(\sigma) = (\operatorname{sgn} \sigma)(L \otimes M)_\sigma.$$

Como L y M son alternadas

$$\psi(\gamma) = L \otimes M$$

para todo γ en G. Luego, como $(N\sigma)\tau = N\tau\sigma$ para cualquier forma $(r + s)$-lineal N sobre V, se tiene

$$\psi(\tau\gamma) = \psi(\tau), \qquad \tau \text{ en } S_{r+s}, \ \gamma \text{ en } G.$$

Esto dice que la aplicación ψ es una constante sobre cada **clase lateral** (a la izquierda) τG del subgrupo G. Si τ_1 y τ_2 están en S_{r+s}, las clases laterales $\tau_1 G$ y $\tau_2 G$ son idénticas o disjuntas. Cada clase lateral tiene $r!s!$ elementos; luego, hay

$$\frac{(r+s)!}{r!s!}$$

clases laterales distintas. Si S_{r+s}/G representa la colección de todas las clases laterales, entonces ψ define una función en S_{r+s}/G; es decir, por lo que se ha visto, hay una función $\tilde{\psi}$ sobre ese conjunto, de modo que

$$\psi(\tau) = \tilde{\psi}(\tau G)$$

para todo τ de S_{r+s}. Si H es una clase lateral a la izquierda de G, entonces $\tilde{\psi}(H) = \psi(\tau)$ para todo τ de H.

Se define ahora el **producto exterior** de las formas multilineales alternadas L y M de grados r y s haciendo

$$L \wedge M = \sum_{H} \tilde{\psi}(H) \tag{5-49}$$

donde H varía sobre S_{r+s}/G. Otro modo de expresar la definición de $L \wedge M$ es la siguiente. Sea S cualquier conjunto de permutaciones de $\{1, \ldots, r+s\}$ que contenga exactamente un elemento de cada clase lateral a la izquierda de G. Entonces

$$L \wedge M = \sum_{\sigma} (\operatorname{sgn} \sigma)(L \otimes M)_\sigma \tag{5-50}$$

donde σ varía sobre S. Claro es que

$$r!s!\, L \wedge M = \pi_{r+s}(L \otimes M)$$

con lo que la nueva definición es equivalente a (5-47) si K es un cuerpo de característica cero.

Teorema 9. *Sea K un anillo conmutativo con unidad y sea V un módulo sobre K. Entonces el producto exterior es una operación asociativa sobre las formas multilineales alternadas sobre V. En otras palabras, si L, M y N son formas multilineales alternadas sobre V de grados r, s y t, respectivamente, entonces*

$$(L \wedge M) \wedge N = L \wedge (M \wedge N).$$

Demostración. Aunque la demostración del Teorema 8 no se aplica aquí, sí sugiere cómo tratar el caso general. Sea $G(r, s, t)$ el subgrupo de S_{r+s+t} que consta de las permutaciones que permutan los conjuntos

$$\{1, \ldots, r\},\ \{r+1, \ldots, r+s\},\ \{r+s+1, \ldots, r+s+t\}$$

sobre sí mismos. Entonces $(\operatorname{sgn} \mu)(L \otimes M \otimes N)_\mu$ es la misma función multilineal para todos los μ de una clase lateral a la izquierda dada de $G(r, s, t)$. Se escoge un elemento de cada clase lateral a la izquierda de $G(r, s, t)$ y sea E la suma de los correspondientes términos $(\operatorname{sgn} \mu)(L \otimes M \otimes N)_\mu$. Entonces E es independiente de cómo se hayan elegido los representantes μ, y

$$r!s!t!\, E = \pi_{r+s+t}(L \otimes M \otimes N).$$

Se verá, ahora, que $(L \wedge M) \wedge N$ y $L \wedge (M \wedge N)$ son iguales a E.

Sea $G(r + s,\ t)$ el subgrupo de S_{r+s+t} que permuta los conjuntos

$$\{1, \ldots, r + s\},\ \{r + s + 1, \ldots, r + s + t\}$$

sobre sí mismos. Sea T cualquier conjunto de permutaciones de $\{1, \ldots, r + s + t\}$ que contenga exactamente un elemento de cada clase lateral de $G(r + s, t)$. Por (5-50)

$$(L \wedge M) \wedge N = \sum_{\tau} (\operatorname{sgn} \tau)[(L \wedge M) \otimes N]_{\tau}$$

donde la suma se extiende sobre las permutaciones τ de T. Ahora bien, sea $G(r,\ s)$ el subgrupo de S_{r+s} que permuta los conjuntos

$$\{1, \ldots, r\},\ \{r + 1, \ldots, r + s\}$$

sobre sí mismos. Sea S cualquier conjunto de permutaciones de $\{1, \ldots, r + s\}$ que contenga exactamente un elemento de cada clase lateral a la izquierda de $G(r,\ s)$. De (5-50) y con lo que se ha visto anteriormente, se sigue que

$$(L \wedge M) \wedge N = \sum_{\sigma,\tau} (\operatorname{sgn} \sigma)(\operatorname{sgn} \tau)[(L \otimes M)_{\sigma} \otimes N]_{\tau}$$

donde la suma se extiende sobre todos los pares σ, τ de $S \times T$. Si se conviene en identificar cada σ de S_{r+s} con el elemento de S_{r+s+t} que coincida con σ en $\{1, \ldots, r + s\}$, y sea la identidad en $\{r + s + 1, \ldots, r + s + t\}$, entonces se puede escribir

$$(L \wedge M) \wedge N = \sum_{\sigma,\tau} \operatorname{sgn} (\sigma \tau)[(L \otimes M \otimes N)_{\sigma}]_{\tau}.$$

Pero

$$[(L \otimes M \otimes N)_{\sigma}]_{\tau} = (L \otimes M \otimes N)_{\tau\sigma}.$$

Luego

$$(L \wedge M) \wedge N = \sum_{\sigma,\tau} \operatorname{sgn} (\tau \sigma)(L \otimes M \otimes N)_{\tau\sigma}.$$

Supóngase ahora que se tiene

$$\tau_1\sigma_1 = \tau_2\sigma_2\gamma$$

con σ_i en S, τ_i en T y γ en $G(r, s, t)$. Entonces $\tau_2^{-1}\tau_1 = \tau_2\gamma\sigma_1^{-1}$ y como $\sigma_2\gamma\sigma_1^{-1}$ está en $G(r + s,\ t)$, se sigue que τ_1 y τ_2 están en la misma clase lateral a la izquierda de $G(r + s,\ t)$. Por tanto, $\tau_1 = \tau_2$ y $\sigma_1 = \sigma_2\gamma$. Pero esto implica que σ_1 y σ_2 (considerados como elementos de S_{r+s}) pertenecen a la misma clase lateral de $G(r,\ s)$; luego $\sigma_1 = \sigma_2$. Por tanto, los productos $\tau\sigma$ que corresponden a los

$$\frac{(r + s + t)!}{(r + s)!t!}\,\frac{(r + s)!}{r!s!}$$

pares $(\tau,\ \sigma)$ de $T \times S$ son todos distintos y están en distintas clases laterales de $G(r,\ s,\ t)$. Como hay exactamente

$$\frac{(r + s + t)!}{r!s!t!}$$

clases laterales a la izquierda de $G(r, s, t)$ en S_{r+s+t}, se sigue que $(L \wedge M) \wedge N = E$. Por razonamiento análogo, también $L \wedge (M \wedge N) = E$. ▌

Ejemplo 13. El producto exterior está íntimamente relacionado con ciertas fórmulas para calcular determinantes, conocidas como **desarrollos de Laplace.** Sea K un anillo conmutativo con unidad y n un entero positivo. Supóngase que $1 \leq r \leq n$, y sea L la forma r-lineal alternada sobre K^n definida por

$$L(\alpha_1, \ldots, \alpha_r) = \det \begin{bmatrix} A_{11} & \cdots & A_{1r} \\ \vdots & & \vdots \\ A_{r1} & \cdots & A_{rr} \end{bmatrix}.$$

Si $s = n - r$, y M es la forma s-lineal alternada

$$M(\alpha_1, \ldots, \alpha_s) = \det \begin{bmatrix} A_{1(r+1)} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{s(r+1)} & \cdots & A_{sn} \end{bmatrix}$$

entonces $L \wedge M = D$, la función determinante sobre K^n. Esto es inmediato, porque $L \wedge M$ es una forma n-lineal alternada y (como puede verse)

$$(L \wedge M)(\epsilon_1, \ldots, \epsilon_n) = 1.$$

Si ahora se describe $L \wedge M$ en forma apropiada, se obtiene un desarrollo de Laplace de la función determinante de una matriz $n \times n$ sobre K.

En el grupo de permutaciones S_n, sea G el subgrupo que permuta los conjuntos $\{1, \ldots, r\}$ y $\{r + 1, \ldots, n\}$ sobre sí mismos. Cada clase lateral a la izquierda de G contiene precisamente una permutación σ tal que $\sigma 1 < \sigma 2 < \cdots < \sigma r$ y $\sigma(r + 1) < \cdots < \sigma n$. El signo de estas permutaciones está dado por

$$\operatorname{sgn} \sigma = (-1)^{\sigma 1 + \cdots + \sigma r + (r(r-1)/2)}.$$

El producto exterior $L \wedge M$ está dado por

$$(L \wedge M)(\alpha_1, \ldots, \alpha_n) = \Sigma\, (\operatorname{sgn} \sigma) L(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma r}) M(\alpha_{\sigma(r+1)}, \ldots, \alpha_{\sigma n})$$

donde la suma se toma sobre una colección de σ, una de cada clase lateral de G. Por tanto,

$$(L \wedge M)(\alpha_1, \ldots, \alpha_n) = \sum_{j_1 < \cdots < j_r} e_J\, L(\alpha_{j_1}, \ldots, \alpha_{j_r}) M(\alpha_{k_1}, \ldots, \alpha_{k_s})$$

donde

$$e_J = (-1)^{j_1 + \cdots + j_r + (r(r-1)/2)}$$
$$k_i = \sigma(r + i).$$

En otras palabras,

$$\det A = \sum_{j_1 < \cdots < j_r} e_J \begin{vmatrix} A_{j_1,1} & \cdots & A_{j_1,r} \\ \vdots & & \vdots \\ A_{j_r,1} & \cdots & A_{j_r,r} \end{vmatrix} \begin{vmatrix} A_{k_1,r+1} & \cdots & A_{k_1,n} \\ \vdots & & \vdots \\ A_{k_s,r+1} & \cdots & A_{k_s,n} \end{vmatrix}$$

Este es un desarrollo de Laplace. Se pueden obtener otros remplazando los

conjuntos $\{1, \ldots, r\}$ y $\{r+1, \ldots, n\}$ por dos conjuntos de índices complementarios diferentes.

Si V es un K-módulo, se pueden reunir las diferentes formas de módulos $\Lambda^r(V)$ y usar el producto exterior para definir un anillo. Por simplificar, se hará esto solo para el caso de un K-módulo libre de rango n. Los módulos $\Lambda^r(V)$ son entonces triviales para $r > n$. Se define

$$\Lambda(V) = \Lambda^0(V) \oplus \Lambda^1(V) \oplus \cdots \oplus \Lambda^n(V).$$

Esta es una suma directa externa —algo de lo que no se ha tratado previamente. Los elementos de $\Lambda(V)$ son los $(n+1)$-tuples $(L_0, \ldots, L_n)$ con L_r en $\Lambda^r(V)$. Adición y multiplicación por elementos de K se definen como se hace para los $(n+1)$-tuples. Incidentalmente, $\Lambda^0(V) = K$. Si se identifica $\Lambda^r(K)$ con los $(n+1)$-tuples $(0, \ldots, 0, L, 0, \ldots, 0)$ donde L está en $\Lambda^r(K)$, entonces $\Lambda^r(K)$ es un submódulo de $\Lambda(V)$ y la descomposición en suma directa

$$\Lambda(V) = \Lambda^0(V) \oplus \cdots \oplus \Lambda^n(V)$$

es válida en el sentido corriente. Como $\Lambda^r(V)$ es un K-módulo de rango $\binom{n}{r}$, se ve que $\Lambda(V)$ es un K-módulo libre y

$$\begin{aligned} \text{rango } \Lambda(V) &= \sum_{r=0}^{n} \binom{n}{r} \\ &= 2^n. \end{aligned}$$

El producto exterior define una multiplicación en $\Lambda(V)$: Utilícese el producto exterior sobre formas y extiéndasele linealmente a $\Lambda(V)$. Este producto es distributivo con respecto a la adición de $\Lambda(V)$ y da a $\Lambda(V)$ estructura de anillo. Este anillo es el **anillo de Grassman** sobre V^*. No es un anillo conmutativo; por ejemplo, si L y M están, respectivamente, en Λ^r y Λ^s, entonces

$$L \wedge M = (-1)^{rs} M \wedge L.$$

El anillo de Grassman es importante en varias partes de la matemática.

6. Formas canónicas elementales

6.1. Introducción

Se ha dicho antes que nuestro objetivo principal es el estudio de las transformaciones lineales sobre un espacio vectorial de dimensión finita. Hasta ahora se han dado ejemplos específicos de transformaciones lineales y se han demostrado algunos teoremas respecto a las transformaciones lineales generales. En el caso de dimensión finita se han empleado bases ordenadas para representar tales transformaciones por medio de matrices; y esta representación ha ayudado a la comprensión de cómo operan. Se ha explorado el espacio vectorial $L(V, W)$, que consta de las transformaciones lineales de un espacio en otro, y el álgebra lineal $L(V, V)$ que consiste en las transformaciones lineales de un espacio en sí mismo.

En los dos capítulos siguientes se estudiarán los operadores lineales. Nuestro programa se propone seleccionar un operador lineal T sobre un espacio de dimensión finita V y «separarlo para ver qué es lo que lo hace importante». En esta primera etapa es más sencillo para nuestro propósito usar el lenguaje matricial. Vale decir, dado un operador lineal T, encontrar una base ordenada de V en la que la matriz de T tome una forma especialmente simple.

He aquí una ilustración de lo que se intenta. Probablemente las matrices más sencillas de manejar, fuera de los múltiplos escalares de la matriz unidad, son las diagonales

(6-1)
$$D = \begin{bmatrix} c_1 & 0 & 0 & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ 0 & 0 & c_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & c_n \end{bmatrix}.$$

Sea T un operador lineal sobre un espacio V de dimensión n. Si se puede encontrar una base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ de V en la que T se pueda expresar por medio de una matriz diagonal D (6-1), podrá tenerse más información respecto a T. Por ejemplo, números simples asociados con T, tales como el rango de T o el determinante de T, pueden determinarse fácilmente por la matriz D. Se pueden expresar explícitamente la imagen y el espacio nulo de T. Como $[T]_{\mathcal{B}} = D$ si, y solo si,

(6-2) $$T\alpha_k = c_k\alpha_k, \qquad k = 1, \ldots, n$$

la imagen será el subespacio generado por aquellos α_k para los que $c_k \neq 0$, y el espacio nulo será generado por los restantes α_k. En efecto, parece justo decir que si se conoce una base $\mathcal{B}$ y una matriz diagonal D para la cual $[T]_{\mathcal{B}} = D$, se podrá responder sin dificultad cualquier pregunta que surja con respecto a T.

¿Se puede representar todo operador lineal T por medio de una matriz diagonal con respecto a alguna base ordenada? Si no, ¿para qué operadores T existe una base semejante? ¿Cómo se puede encontrar tal base si existe? Si tal base no existe, ¿cuál es el tipo más sencillo de matriz por la cual puede representarse T? Estas son algunas de las preguntas a que se debe responder en este (y en el siguiente) capítulo. El tratamiento se hará más complicado a medida que se vaya viendo cuáles son las dificultades.

6.2. *Valores propios*

Las observaciones de la introducción sugieren un punto de partida para el análisis del operador lineal general T. Se comienza con (6-2), que sugiere se estudien los vectores que son transformados por T en múltiplos de sí mismos.

Definición *Sea V un espacio vectorial sobre el cuerpo F y sea T un operador lineal sobre V. Un* **valor propio** *de T es un escalar c de F tal que existe un vector no nulo α con $T\alpha = c\alpha$. Si c es un valor propio de T, entonces*

(a) *cualquier α tal que $T\alpha = c\alpha$ se llama un* **vector propio** *de T asociado al valor propio c;*

(b) *la colección de todos los α tales que $T\alpha = c\alpha$ se llama espacio* **propio asociado** *a c.*

Los valores propios se llaman también a menudo raíces características, eigenvalores, valores característicos o valores espectrales. En este libro se usará solo el nombre de «valores propios».

Si T es cualquier operador y c es cualquier escalar, el conjunto de los vectores α tales que $T\alpha = c\alpha$ es un subespacio de V. Es el espacio nulo de la transformación lineal $(T - cI)$. Se llama a c un valor propio de T si este subespacio es distinto del subespacio nulo, es decir, si $(T - cI)$ no es inyectiva. Si el espacio soporte V es de dimensión finita $(T - cI)$, no es inyectiva justamente cuando su determinante es distinto de 0. En resumen:

Teorema 1. *Sea T un operador lineal sobre un espacio V de dimensión finita y sea c un escalar. Las siguientes afirmaciones son equivalentes.*

(i) *c es un valor propio de T.*
(ii) *El operador $(T - cI)$ es singular (no inversible).*
(iii) *det $(T - cI) = 0$.*

El criterio del determinante (iii) es muy importante, porque dice cómo se pueden encontrar los valores propios de T. Como det $(T - cI)$ es un polinomio de grado n en la variable c, se determinarán los valores propios como las raíces de tal polinomio. Se explica esto con más detalle.

Si $\mathcal{B}$ es cualquier base ordenada de V y $A = [T]_{\mathcal{B}}$, entonces $(T - cI)$ es inversible si, y solo si, la matriz $(A - cI)$ es inversible. En consecuencia, se tiene la siguiente definición.

Definición. *Si A es una matriz $n \times n$ sobre el cuerpo F, un* **valor propio** *de A en F es un escalar c de F tal que la matriz $(A - cI)$ es singular (no inversible).*

Como c es un valor propio de A si, y solo si, det $(A - cI) = 0$, o en forma equivalente, si, y solo si, det $(cI - A) = 0$, se puede construir la matriz $(xI - A)$ con elementos polinómicos y considerar el polinomio $f = \det(xI - A)$. En tal caso los valores propios de A en F son los escalares c en F tales que $f(c) = 0$. Por esta razón a f se le llama el **polinomio característico** de A. Es importante observar que f es un polinomio mónico de grado n. Lo cual es fácilmente comprobable por la fórmula para el determinante de una matriz en términos de sus elementos.

Lema. *Las matrices semejantes tienen el mismo polinomio característico.*

Demostración. Si $B = P^{-1}AP$, entonces

$$\begin{aligned}\det(xI - B) &= \det(xI - P^{-1}AP)\\ &= \det(P^{-1}(xI - A)P)\\ &= \det P^{-1} \cdot \det(xI - A) \cdot \det P\\ &= \det(xI - A). \quad \blacksquare\end{aligned}$$

Este lema permite definir el polinomio característico de un operador T como el polinomio característico de cualquier matriz $n \times n$ que representa a T en alguna base ordenada de V. Al igual que para las matrices, los valores propios de T serán las raíces del polinomio característico de T. En particular, esto muestra que T no puede tener más de n valores propios distintos. Es importante señalar que T puede carecer de valores propios.

Ejemplo 1. Sea T el operador lineal sobre R^2 representado, en la base canónica, por la matriz

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

El polinomio característico de T (o de A) es

$$\det(xI - A) = \begin{vmatrix} x & 1 \\ -1 & x \end{vmatrix} = x^2 + 1.$$

Como este polinomio tiene dos raíces no reales, T no tiene valores propios. Si U es el operador lineal definido sobre C^2 y representado por A en la base ordenada canónica, entonces U sí tiene dos valores propios, i y $-i$ en C.

Ejemplo 2. Sea A la matriz (real) 3×3

$$\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}.$$

Entonces el polinomio característico de A es

$$\begin{vmatrix} x-3 & -1 & 1 \\ -2 & x-2 & 1 \\ -2 & -2 & x \end{vmatrix} = x^3 - 5x^2 + 8x - 4 = (x-1)(x-2)^2.$$

Con lo que los valores propios de A son 1 y 2.

Supóngase que T es el operador lineal sobre R^3 representado por A en la base canónica. Hállense los vectores propios de T asociados a los valores propios 1 y 2. Ahora

$$A - I = \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix}.$$

Es inmediatamente obvio que $A - I$ tiene rango 2 (y así, pues, $T - I$ tiene nulidad 1). Con lo que el espacio de los vectores propios asociado al valor propio 1 es unidimensional. El vector $\alpha_1 = (1, 0, 2)$ genera el espacio nulo de $T - I$. Así que $T\alpha = \alpha$ si, y solo si, α es un múltiplo escalar de α_1. Sea ahora

$$A - 2I = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix}.$$

Evidentemente, $A - 2I$ también tiene rango 2, así que el espacio de los vectores propios asociado al valor propio 2 tiene dimensión 1. Es evidente que $T\alpha = 2\alpha$ si, y solo si, α es un múltiplo escalar de $\alpha_2 = (1, 1, 2)$.

Definición *Sea T un operador lineal sobre un espacio vectorial de dimensión finita V. Se dice que T es* **diagonalizable** *si existe una base de V tal que cada vector suyo sea vector propio de T.*

La razón del nombre es clara; en efecto, si existe una base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ de V en la que cada α_i es un vector propio de T, entonces la matriz de T en la base ordenada $\mathcal{B}$ es diagonal. Si $T\alpha_i = c_i\alpha_i$, entonces

$$[T]_{\mathcal{B}} = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix}.$$

Por cierto no se requiere que los escalares $c_1, \ldots, c_n$ sean distintos, incluso pueden ser todos iguales (cuando T es un múltiplo escalar del operador identidad).

Se podría también definir T como diagonalizable cuando los vectores propios de T generan V. Esto es solo en apariencia distinto a la definición dada, ya que se puede formar una base tomándola de cualquier conjunto generador de vectores.

En los Ejemplos 1 y 2 se han elegido a propósito operadores lineales T sobre R^n que no son diagonalizables. En el Ejemplo 1 se tiene un operador lineal sobre R^2 que no es diagonalizable, porque no tiene valores propios. En el Ejemplo 2 el operador lineal T tiene valores propios; en efecto, el polinomio característico de T se pudo factorizar completamente sobre el cuerpo de los números reales: $f = (x - 1)(x - 2)^2$. Sin embargo, T no es diagonalizable, pues existe solo un espacio unidimensional de vectores propios asociados a cada valor propio de T. Por tanto, no es posible formar una base de R^3 con vectores propios de T.

Supóngase que T es un operador lineal diagonalizable. Sean $c_1, \ldots, c_k$ los valores propios *distintos* de T. Entonces existe una base ordenada $\mathcal{B}$ con respecto a la cual T está representado por una matriz diagonal; es decir, los elementos de la diagonal son los escalares c_i, cada uno de los cuales se repite un cierto número de veces. Si c_i aparece repetida d_i veces, entonces (se puede disponer así) la matriz tiene la forma de bloque

$$(6\text{-}3) \qquad [T]_{\mathcal{B}} = \begin{bmatrix} c_1 I_1 & 0 & \cdots & 0 \\ 0 & c_2 I_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_k I_k \end{bmatrix}$$

donde I_j es la matriz unidad $d_j \times d_j$. De esta matriz se observan dos cosas. Primero, el polinomio característico de T es producto de factores lineales (posiblemente repetidos):

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}.$$

Si el cuerpo escalar F es algebraicamente cerrado, v.g., el cuerpo de los números complejos, todo polinomio sobre F puede ser factorizado (véase Sección 4.5); sin embargo, si F no es algebraicamente cerrado, se tendrá una propiedad especial de T cuando se diga que sus polinomios característicos tienen tal facto-

rización. Lo segundo que se observa de (6-3) es que d_i, el número de veces que c_i se repite como raíz de f, es igual a la dimensión del espacio de los vectores propios asociados al valor propio c_i. Ello es así porque la nulidad de una matriz diagonal es igual al número de ceros que tiene en la diagonal principal, y la matriz $[T - c_iI]_{\mathcal{B}}$ tiene d_i ceros en su diagonal principal. Esta relación entre la dimensión del espacio propio y la multiplicidad de los valores propios, como una raíz de f, no parece ser importante al comienzo; sin embargo, proveerá de un medio simple para determinar cuándo un operador dado es diagonalizable.

Lema. *Supóngase que* $T\alpha = c\alpha$. *Si* f *es cualquier polinomio, entonces* $f(T)\alpha = f(c)\alpha$.

Demostración. Se deja como ejercicio.

Lema. *Sea* T *un operador lineal sobre un espacio vectorial de dimensión finita. Sean* $c_1, \ldots, c_k$ *los valores propios distintos de* T, *y sea* W_i *el espacio de los vectores propios asociados con el valor característico* c_i. *Si* $W = W_1 + \cdots + W_k$, *entonces*

$$\dim W = \dim W_1 + \cdots + \dim W_k.$$

En efecto, si $\mathcal{B}_i$ *es una base ordenada de* W_i, *entonces* $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ *es una base ordenada de* W.

Demostración. El espacio $W = W_1 + \cdots + W_k$ es el subespacio generado por todos los vectores propios de T. Normalmente, cuando se forma la suma W de los subespacios W_i, se expresa que $\dim W < \dim W_1 + \cdots + \dim W_k$, por las relaciones lineales que pueden existir entre los vectores de los diferentes espacios. Este lema afirma que los espacios propios asociados a los diferentes valores propios son independientes uno de otro.

Supóngase que (para cada i) se tenga un vector β_i en W_i y supóngase que $\beta_1 + \cdots + \beta_k = 0$. Se demostrará que $\beta_i = 0$ para cada i. Sea f un polinomio. Como $T\beta_i = c_i\beta_i$, el lema anterior dice que

$$\begin{aligned} 0 = f(T)0 &= f(T)\beta_1 + \cdots + f(T)\beta_k \\ &= f(c_1)\beta_1 + \cdots + f(c_k)\beta_k. \end{aligned}$$

Eligiendo los polinomios $f_1, \ldots, f_k$ de modo que

$$f_i(c_j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

Entonces

$$\begin{aligned} 0 = f_i(T)0 &= \sum_j \delta_{ij}\beta_j \\ &= \beta_i. \end{aligned}$$

Ahora sea $\mathcal{B}_i$ una base ordenada de W_i y sea $\mathcal{B}$ la sucesión $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$. Entonces $\mathcal{B}$ genera el subespacio $W = W_1 + \cdots + W_k$. También $\mathcal{B}$ es una

sucesión linealmente independiente de vectores, por lo siguiente. Cualquier relación lineal entre los vectores en $\mathcal{B}$ tendrá la forma $\beta_1 + \cdots + \beta_k = 0$, donde β_i es alguna combinación lineal de vectores en $\mathcal{B}_i$. De lo anterior se sabe que $\beta_i = 0$ para cada i. Como cada $\mathcal{B}_i$ es linealmente independiente, se ve que solo se tiene la relación lineal trivial entre los vectores en $\mathcal{B}$. ∎

Teorema 2. *Sea T un operador lineal sobre un espacio vectorial V de dimensión finita. Sean $c_1, \ldots, c_k$ los valores propios distintos de T y sean W_i el espacio nulo de $(T - c_iI)$. Lo siguiente es equivalente.*

(i) *T es diagonalizable.*

(ii) *El polinomio característico de T es*

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

y dim $W_i = d_i$, $i = 1, \ldots, k$.

(iii) *dim $W_1 + \cdots +$ dim $W_k =$ dim V.*

Demostración. Ya se observó que (i) implica a (ii). Si el polinomio característico f es producto de factores lineales, como en (ii), entonces $d_1 + \cdots + d_k =$ dim V. En efecto, la suma de los d_i es el grado del polinomio característico y ese grado es dim V. Por tanto, (ii) implica (iii). Supóngase que se tenga (iii). Por el lema, se debe tener que $V = W_1 + \cdots + W_k$, es decir, que los vectores propios de T generan V. ∎

El análogo al Teorema 2 para matrices puede formularse en la siguiente forma. Sea A una matriz $n \times n$ con elementos en un cuerpo F y sean $c_1, \ldots, c_k$ los valores propios distintos de A en F. Para cada i, sea W_i el espacio de las matrices columnas X (con elementos en F) tales que

$$(A - c_iI)X = 0,$$

y sea $\mathcal{B}_i$ una base ordenada de W_i. Las bases $\mathcal{B}_1, \ldots, \mathcal{B}_k$ se juntan para formar la sucesión de columnas de una matriz P:

$$P = [P_1, P_2, \ldots] = (\mathcal{B}_1, \ldots, \mathcal{B}_k).$$

La matriz A es semejante sobre F a una matriz diagonal si, y solo si, P es una matriz cuadrada. Cuando P es cuadrada, P es inversible y $P^{-1}AP$ es diagonal.

Ejemplo 3. Sea T un operador lineal sobre R^3 representado en la base ordenada canónica por la matriz

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}.$$

Se indica cómo se puede calcular el polinomio característico usando sucesivas operaciones de fila y columna:

$$\begin{vmatrix} x-5 & 6 & 6 \\ 1 & x-4 & -2 \\ -3 & 6 & x+4 \end{vmatrix} = \begin{vmatrix} x-5 & 0 & 6 \\ 1 & x-2 & -2 \\ -3 & 2-x & x+4 \end{vmatrix}$$

$$= (x-2)\begin{vmatrix} x-5 & 0 & 6 \\ 1 & 1 & -2 \\ -3 & -1 & x+4 \end{vmatrix}$$

$$= (x-2)\begin{vmatrix} x-5 & 0 & 6 \\ 1 & 1 & -2 \\ -2 & 0 & x+2 \end{vmatrix}$$

$$= (x-2)\begin{vmatrix} x-5 & 6 \\ -2 & x+2 \end{vmatrix}$$

$$= (x-2)(x^2-3x+2)$$
$$= (x-2)^2(x-1).$$

¿Cuáles son las dimensiones de los espacios de los vectores propios asociados con los dos valores propios? Se tiene

$$A - I = \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix}$$

$$A - 2I = \begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix}.$$

Se sabe que $A - I$ es singular y es claro que rango $(A - I) \geq 2$. Por tanto, rango $(A - I) = 2$. Es evidente que rango $(A - 2I) = 1$.

Sean W_1 y W_2 los espacios de los vectores propios asociados con los valores propios 1 y 2. Se sabe que dim $W_1 = 1$ y que dim $W_2 = 2$. Por el Teorema 2, T es diagonalizable. Es fácil obtener una base para R^3 en que T esté representado por una matriz diagonal. El espacio nulo de $(T - I)$ es generado por el vector $\alpha_1 = (3, -1, 3)$, con lo que $\{\alpha_1\}$ es una base para W_1. El espacio nulo de $T - 2I$ (es decir, el espacio W_2) consta de los vectores (x_1, x_2, x_3) con $x_1 = 2x_2 + 2x_3$. Así, un ejemplo de una base de W_2 es

$$\alpha_2 = (2, 1, 0)$$
$$\alpha_3 = (2, 0, 1).$$

Si $\mathcal{B} = \{\alpha_1, \alpha_2, \alpha_3\}$, entonces $[T]_{\mathcal{B}}$ es la matriz diagonal

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Que T sea diagonalizable quiere decir que la matriz original A es semejante (sobre R) a la matriz diagonal D. La matriz P, que permite cambiar las coordenadas de la base $\mathcal{B}$ a la base canónica, es (claramente) la matriz que tiene las transpuestas de α_1, α_2, α_3 como vectores columnas:

$$P = \begin{bmatrix} 3 & 2 & 2 \\ -1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}.$$

además, $AP = PD$, con lo que

$$P^{-1}AP = D.$$

Ejercicios

1. En cada uno de los siguientes casos, sea T el operador lineal sobre R^2 representado por la matriz A en la base ordenada canónica de R^2 y sea U el operador lineal en C^2 representado por A en la base ordenada canónica. Encontrar el polinomio característico de T y de U, hallar los valores propios de cada operador y para cada uno de tales valores propios c hallar una base para el correspondiente espacio de vectores característicos.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 2 & 3 \\ -1 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

2. Sea V un espacio vectorial de dimensión n sobre F. ¿Cuál es el polinomio característico del operador identidad sobre V? ¿Cuál es el polinomio característico para el operador cero?

3. Sea A una matriz triangular $n \times n$ sobre el cuerpo F. Demostrar que los valores propios de A son los elementos de la diagonal de A, es decir, los escalares A_{ii}.

4. Sea T un operador lineal sobre R^3 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}.$$

Demostrar que T es diagonalizable construyendo una base para R^3, cada vector de la cual es un vector propio de T.

5. Sea

$$A = \begin{bmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{bmatrix}.$$

¿Es A semejante, sobre el cuerpo R, a una matriz diagonal? ¿Es A semejante, sobre el cuerpo C, a una matriz diagonal?

6. Sea T el operador lineal sobre R^4 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \end{bmatrix}.$$

¿En qué condiciones para a, b y c es T diagonalizable?

7. Sea T un operador lineal sobre el espacio vectorial V de dimensión n y supóngase que T tiene n valores propios *distintos*. Demostrar que T es diagonalizable.

8. Sean A y B matrices $n \times n$ sobre el cuerpo F. Demostrar que si $(I - AB)$ es inversible, entonces $I - BA$ es inversible y que

$$(I - BA)^{-1} = I + B(I - AB)^{-1}A.$$

9. Usar el resultado del Ejercicio 8 para demostrar que si A y B son matrices $n \times n$ sobre el cuerpo F, entonces AB y BA tienen los mismos valores propios en F.

10. Supóngase que A es una matriz 2×2 con elementos reales, simétrica ($A^t = A$). Demostrar que A es semejante, sobre R, a una matriz diagonal.

11. Sea N una matriz compleja 2×2 tal que $N^2 = 0$. Demostrar que $N = 0$ o N es semejante, sobre C, a

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

12. Usar el resultado del Ejercicio 11 para demostrar lo siguiente: Si A es una matriz 2×2 de elementos complejos, entonces A es semejante, sobre C, a una matriz de uno de los dos tipos siguientes:

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \qquad \begin{bmatrix} a & 0 \\ 1 & a \end{bmatrix}.$$

13. Sea V el espacio vectorial de las funciones continuas de R en R; es decir, el espacio de las funciones continuas de valor real en el eje real. Sea T el operador lineal sobre V definido por

$$(Tf)(x) = \int_0^x f(t)\, dt.$$

Demostrar que T no tiene valores propios.

14. Sea A una matriz *diagonal* $n \times n$ de polinomio característico

$$(x - c_1)^{d_1} \cdots (x - c_k)^{d_k},$$

donde los $c_1, \ldots, c_k$ son distintos. Sea V el espacio de las matrices $n \times n$, B, tales que $AB = BA$. Demostrar que la dimensión de V es $d_1^2 + \cdots + d_k^2$.

15. Sea V el espacio de las matrices $n \times n$ sobre F. Sea A una matriz $n \times n$ dada sobre F. Sea T el operador lineal «multiplicación a la izquierda por A» en V. ¿Tienen A y T los mismos valores propios?

6.3. *Polinomios anuladores*

Cuando se procura analizar un operador lineal T, una de las cosas más útiles de conocer es la clase de los polinomios que anulan a T. Para precisar, supóngase que T es un operador lineal sobre V, espacio vectorial sobre el cuerpo F. Si p es un polinomio sobre F, entonces $p(T)$ es también un operador lineal sobre V. Si q es otro polinomio sobre F, entonces

$$(p + q)(T) = p(T) + q(T)$$
$$(pq)(T) = p(T)q(T).$$

Por tanto, la colección de polinomios p que anulan a T, en el sentido de que

$$p(T) = 0,$$

es un ideal en el álgebra de los polinomios $F[x]$. Puede ser el ideal cero; es decir. puede ser que T no sea anulado por cualquier polinomio no nulo. Pero ello no puede suceder si el espacio V es de dimensión finita.

Supóngase que T es un operador lineal sobre el espacio V de dimensión n. Considérense las primeras $(n^2 + 1)$ potencias de T:

$$I, T, T^2, \ldots, T^{n^2}.$$

Esta es una sucesión de $n^2 + 1$ operadores en $L(V, V)$, el espacio de los operadores lineales sobre V. El espacio $L(V, V)$ tiene dimensión n^2. Por tanto, la sucesión de los $n^2 + 1$ operadores debe ser linealmente dependiente, es decir, se debe tener que

$$c_0 I + c_1 T + \cdots + c_{n^2} T^{n^2} = 0$$

para los escalares c_i, no todos nulos. En consecuencia, el ideal de polinomios que anulan a T tiene un polinomio no nulo de grado n^2 o menor.

Conforme al Teorema 5 del Capítulo 4 todo ideal de polinomios consta de todos los múltiplos de un cierto polinomio mónico fijo, el generador del ideal. Así, pues, al operador T corresponde un múltiplo mónico p con la siguiente propiedad: Si f es un polinomio sobre F, entonces $f(T) = 0$ si, y solo si, $f = pg$, donde g es algún polinomio sobre F.

Definición. *Sea T un operador lineal sobre un espacio vectorial V de dimensión finita sobre el cuerpo F. El* **polinomio minimal** *de T es el polinomio mónico generador (único) del ideal de polinomios sobre F que anulan a T.*

El nombre de «polinomio minimal» procede de que el generador de un ideal de polinomios está caracterizado por ser el polinomio mónico de grado mínimo en el ideal. Lo cual quiere decir que el polinomio minimal p para el operador lineal T está unívocamente determinado por estas tres propiedades:

(1) p es un polinomio mónico sobre el cuerpo escalar.
(2) $p(T) = 0$.
(3) Ningún polinomio sobre F que anule a T tiene grado menor que el que tiene p.

Si A es una matriz $n \times n$ sobre F, se define el **polinomio minimal** de A en forma análoga, como el único generador mónico del ideal de todos los polinomios sobre F que anulan a A. Si el operador T está representado en cierta base ordenada por la matriz A, entonces T y A tienen el mismo polinomio minimal. Esto es porque $f(T)$ está representado en la base por la matriz $f(A)$, de modo que $f(T) = 0$ si, y solo si, $f(A) = 0$.

De la última observación respecto a operadores y matrices se sigue que las

matrices semejantes tienen el mismo polinomio minimal. Ese hecho resulta también evidente de la definición, ya que

$$f(P^{-1}AP) = P^{-1}f(A)P$$

para todo polinomio f.

Queremos hacer otra observación respecto a los polinomios minimales de las matrices. Supóngase que A es una matriz $n \times n$ con elementos en el cuerpo F. Supóngase que F_1 es un cuerpo que contiene a F como subcuerpo. (Por ejemplo, A puede ser una matriz con elementos racionales, mientras que F_1 es el cuerpo de los números reales. O A puede ser una matriz con elementos reales, mientras que F_1 es el cuerpo de los números complejos.) Se puede considerar A como matriz $n \times n$ sobre F o como matriz $n \times n$ sobre F_1. A primera vista podría creerse que se obtienen dos polinomios minimales distintos para A. Por suerte, ese no es el caso; y debemos ver por qué. ¿Cuál es la definición del polinomio minimal para A, considerándola como una matriz $n \times n$ sobre el cuerpo F? Consideramos todos los polinomios mónicos, con coeficientes en F, que anulan a A y tomamos el de menor grado. Si f es un polinomio mónico sobre F:

$$f = x^k + \sum_{j=0}^{k-1} a_j x^j \tag{6-4}$$

entonces $f(A) = 0$ dice simplemente que tenemos una relación lineal entre potencias de A:

$$A^k + a_{k-1}A^{k-1} + \cdots + a_1A + a_0I = 0. \tag{6-5}$$

El grado del polinomio minimal es el menor entero positivo k tal que existe una relación lineal de la forma (6-5) entre las potencias de $I, A, \ldots, A^k$. Más aún, por la unicidad del polinomio minimal, existe para aquel k una, y solo una, relación de la forma (6-5); es decir, una vez que el minimal k está determinado, existen escalares únicos $a_0, \ldots, a_{k-1}$ en F para los que es válido (6-5). Son los coeficientes del polinomio minimal.

Ahora (para cada k) se tiene en (6-5) un sistema de n^2 ecuaciones lineales para las «incógnitas» $a_0, \ldots, a_{k-1}$. Como los elementos de A están en F, los coeficientes del sistema de ecuaciones (6-5) están en F. Por tanto, si el sistema tiene una solución con los $a_0, \ldots, a_{k-1}$ en F_1, tiene una solución con los $a_0, \ldots, a_{k-1}$ en F. (Véase el final de la Sección 1.4.) Está claro ahora que los dos polinomios minimales son los mismos.

¿Qué es lo que se sabe hasta ahora con respecto al polinomio minimal de un operador lineal en un espacio de dimensión n? Solo que su grado no excede de n^2. Pero ésta es una estimación poco aproximada, ya que el grado no puede exceder a n. Se demostrará en breve que el operador es anulado por su polinomio característico. Pero antes queremos observar algo más elemental.

Teorema 3. *Sea T un operador lineal sobre un espacio vectorial V de dimensión n (o sea A una matriz $n \times n$). El polinomio característico y el polinomio minimal de T (de A) tienen las mismas raíces, salvo multiplicidades.*

Demostración. Sea p el polinomio minimal de T. Sea c un escalar. Lo que se desea demostrar es que $p(c) = 0$ si, y solo si, c es un valor propio de T.

Primero supóngase que $p(c) = 0$. Entonces

$$p = (x - c)q$$

donde q es un polinomio. Como grd $q <$ grd p, la definición del polinomio minimal p dice que $q(T) \neq 0$. Se elige un vector β tal que $q(T)\beta \neq 0$. Sea $\alpha = q(T)\beta$. Entonces

$$\begin{aligned} 0 &= p(T)\beta \\ &= (T - cI)q(T)\beta \\ &= (T - cI)\alpha \end{aligned}$$

y así c es un valor propio de T.

Supóngase ahora que c es un valor propio de T, o sea $T\alpha = c\alpha$, con $\alpha \neq 0$. Como se observó en el lema anterior,

$$p(T)\alpha = p(c)\alpha.$$

Ya que $p(T) = 0$ y $\alpha \neq 0$, se tiene que $p(c) = 0$. ▌

Sea T un operador lineal diagonalizable y sean $c_1, \ldots, c_k$ los valores propios distintos de T. Entonces es fácil ver que el polinomio minimal de T es el polinomio

$$p = (x - c_1) \cdots (x - c_k).$$

Si α es un vector propio, entonces uno de los operadores $T - c_1I, \ldots, T - c_kI$ aplica α en 0. Por tanto,

$$(T - c_1I) \cdots (T - c_kI)\alpha = 0$$

para todo vector propio α. Existe una base para el espacio soporte que consta de los vectores propios de T; luego

$$p(T) = (T - c_1I) \cdots (T - c_kI) = 0.$$

Lo que concluimos es lo siguiente. Si T es un operador lineal diagonalizable, entonces el polinomio minimal de T es un producto de factores lineales distintos. Como se verá más adelante, esta propiedad caracteriza los operadores diagonalizables.

Ejemplo 4. Se desea encontrar el polinomio minimal para los operadores en los Ejemplos 1, 2 y 3. Se analizarán en orden inverso. El operador del Ejemplo 3 era diagonalizable con polinomio característico

$$f = (x - 1)(x - 2)^2.$$

Por el párrafo anterior se sabe que el polinomio minimal de T es

$$p = (x - 1)(x - 2).$$

El lector podrá verificar directamente que

$$(A - I)(A - 2I) = 0.$$

En el Ejemplo 2 el operador T también tiene el polinomio característico $f = (x - 1)(x - 2)^2$. Pero este T no es diagonalizable, por lo que no se puede saber que el polinomio minimal sea $(x - 1)(x - 2)$. ¿Qué es lo que se sabe con respecto al polinomio minimal en este caso? Por el Teorema 3 sabemos que sus raíces son 1 y 2, permitiéndose algunas multiplicidades. Luego, p se debe buscar entre los polinomios de la forma $(x - 1)^k(x - 2)^l$, $k \geq 1$, $l \geq 1$. Probar $(x - 1)(x - 2)$:

$$(A - I)(A - 2I) = \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -1 \\ 2 & 0 & -1 \\ 4 & 0 & -2 \end{bmatrix}.$$

Con lo que el polinomio minimal tiene a lo menos grado 3. Luego se puede probar $(x - 1)^2(x - 2)$ o $(x - 1)(x - 2)^2$. Como el segundo es el polinomio característico, parece ser una elección menos al azar. Se puede en realidad calcular que $(A - I)(A - 2I)^2 = 0$. Con lo que el polinomio minimal de T es su polinomio característico.

En el Ejemplo 1 se estudió el operador lineal T sobre R^2 representado en la base canónica por la matriz

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

El polinomio característico es $x^2 + 1$, que no tiene raíces reales. Para determinar el polinomio minimal hay que olvidarse de T y atender a A. Como una matriz compleja 2×2, A tiene los valores propios i y $-i$. Ambas raíces deben figurar en el polinomio minimal. Así el polinomio minimal es divisible por $x^2 + 1$. Es trivial verificar que $A^2 + I = 0$. Con lo que el polinomio minimal es $x^2 + 1$.

Teorema 4 (Cayley-Hamilton). *Sea T un operador lineal sobre un espacio vectorial V de dimensión finita. Si f es el polinomio característico de T, entonces $f(T) = 0$; es decir, el polinomio minimal divide al polinomio característico de T.*

Demostración. Más adelante se darán dos demostraciones más de este teorema, independientemente de la que se dará ahora. La presente demostración, aunque breve, puede ser difícil de entender. Aparte de lo breve, tiene la virtud de dar una clara y nada trivial aplicación de la teoría general de los determinantes desarrollada en el Capítulo 5.

Sea K el anillo conmutativo con unidad que consta de todos los polinomios en T. Es claro que K es un álgebra conmutativa con unidad sobre el cuerpo

escalar. Elijamos una base ordenada $\{\alpha_1, \ldots, \alpha_n\}$ para V, y sea A la matriz que representa a T en esa base dada. Entonces

$$T\alpha_i = \sum_{j=1}^{n} A_{ji}\alpha_j, \qquad 1 \leq i \leq n.$$

Estas ecuaciones pueden escribirse en la forma equivalente

$$\sum_{j=1}^{n} (\delta_{ij}T - A_{ji}I)\alpha_j = 0, \qquad 1 \leq i \leq n.$$

Sea B el elemento de $K^{n\times n}$ con elementos

$$B_{ij} = \delta_{ij}T - A_{ji}I.$$

Cuando $n = 2$

$$B = \begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix}$$

y

$$\begin{aligned} \det B &= (T - A_{11}I)(T - A_{22}I) - A_{12}A_{21}I \\ &= T^2 - (A_{11} + A_{22})T + (A_{11}A_{22} - A_{12}A_{21})I \\ &= f(T) \end{aligned}$$

donde f es el polinomio característico:

$$f = x^2 - (\text{traza } A)x + \det A.$$

Para el caso $n > 2$, es también claro que

$$\det B = f(T)$$

ya que f es el determinante de la matriz $xI - A$ cuyos elementos son los polinomios

$$(xI - A)_{ij} = \delta_{ij}x - A_{ji}.$$

Queremos demostrar que $f(T) = 0$. Para que $f(T)$ sea el operador cero es necesario y suficiente que $(\det B)\alpha_k = 0$ para $k = 1, \ldots, n$. Por la definición de B, los vectores $\alpha_1, \ldots, \alpha_n$ satisfacen las ecuaciones

$$\text{(6-6)} \qquad \sum_{j=1}^{n} B_{ij}\alpha_j = 0, \qquad 1 \leq i \leq n.$$

Cuando $n = 2$, se sugiere escribir (6-6) en la forma

$$\begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

En este caso, la adjunta, adj B, es la matriz

$$\tilde{B} = \begin{bmatrix} T - A_{22}I & A_{21}I \\ A_{12}I & T - A_{11}I \end{bmatrix}$$

y

$$\tilde{B}B = \begin{bmatrix} \det B & 0 \\ 0 & \det B \end{bmatrix}.$$

Luego, se tiene

$$\begin{aligned}(\det B)\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} &= (\tilde{B}B)\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}\\ &= \tilde{B}\left(B\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}\right)\\ &= \begin{bmatrix}0\\ 0\end{bmatrix}.\end{aligned}$$

En el caso general, sea $\tilde{B} = \operatorname{adj} B$. Entonces por (6-6)

$$\sum_{j=1}^{n} \tilde{B}_{ki}B_{ij}\alpha_j = 0$$

para todo par k, i, y sumando sobre i, se tiene

$$\begin{aligned}0 &= \sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{B}_{ki}B_{ij}\alpha_j\\ &= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} \tilde{B}_{ki}B_{ij}\right)\alpha_j.\end{aligned}$$

Ahora $\tilde{B}B = (\det B)I$, con lo que

$$\sum_{i=1}^{n} \tilde{B}_{ki}B_{ij} = \delta_{kj}\det B.$$

Por tanto,

$$\begin{aligned}0 &= \sum_{j=1}^{n} \delta_{kj}(\det B)\alpha_j\\ &= (\det B)\alpha_k, \qquad 1 \le k \le n. \quad \blacksquare\end{aligned}$$

El teorema de Cayley-Hamilton es útil en este momento, fundamentalmente porque reduce la búsqueda del polinomio minimal de varios operadores. Si se conoce la matriz A que representa T en cierta base ordenada, entonces se puede calcular el polinomio característico f. Se sabe que el polinomio minimal p divide a f y que los dos polinomios tienen las mismas raíces. No existe método para calcular en forma precisa las raíces de un polinomio (a menos que su grado sea bajo); sin embargo, si f está factorizado como

(6-7) $\quad f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}, \qquad c_1, \ldots, c_k$ diferentes, $d_i \ge 1$

entonces

(6-8) $\quad p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}; \qquad 1 \le r_j \le d_j.$

Esto es todo lo que se puede decir en general. Si f es el polinomio (6-7) y tiene grado n, entonces para cada polinomio p, como en (6-8), se puede encontrar una matriz $n \times n$ que tiene f como un polinomio característico, y p es su polinomio minimal. No se demostrará esto ahora. Pero queremos destacar el hecho

de que el conocimiento que el polinomio característico tiene la forma (6-7), nos dice que el polinomio minimal tiene la forma (6-8), y nada más nos dice con respecto a p.

Ejemplo 5. Sea A la matriz (racional) 4×4

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}.$$

Las potencias de A son fácilmente calculables:

$$A^2 = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \end{bmatrix}.$$

Así, $A^3 = 4A$, es decir, si $p = x^3 = 4x = x(x + 2)(x - 2)$, entonces $p(A) = 0$. El polinomio minimal, obviamente, no es de grado 1, ya que ello querría decir que A es un múltiplo escalar de la unidad. Luego los posibles polinomios minimales son: p, $x(x + 2)$, $x(x - 2)$, $x^2 - 4$. Los tres polinomios cuadráticos pueden ser eliminados, ya que es inmediatamente obvio que $A^2 \neq -2A$, $A^2 \neq 2A$, $A^2 \neq 4I$. Por tanto, p es el polinomio minimal de A. En particular, 0, 2 y -2 son los valores propios de A. Uno de los factores x, $x - 2$, $x + 2$, debe repetirse dos veces en el polinomio característico. Evidentemente, rango $(A) = 2$. En consecuencia, existe un espacio de vectores propios de dimensión dos asociado al valor característico 0. Por el Teorema 2 queda ahora claro que el polinomio característico es $x^2(x^2 - 4)$ y que A es semejante a la matriz sobre el cuerpo de los números racionales.

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}.$$

Ejercicios

1. Sea V un espacio vectorial de dimensión finita. ¿Cuál es el polinomio minimal para el operador identidad sobre V? ¿Cuál es el polinomio minimal para el operador cero?

2. Sean a, b y c elementos de un cuerpo F y sea A la siguiente matriz 3×3 sobre F:

$$A = \begin{bmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{bmatrix}.$$

Demostrar que el polinomio característico de A es $x^3 - ax^2 - bx - c$ y que éste es también el polinomio minimal de A.

3. Sea A la matriz real 4×4:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & -2 & 2 & 1 \\ 1 & 1 & -1 & 0 \end{bmatrix}.$$

Demostrar que el polinomio característico de A es $x^2(x - 1)^2$ y que es al propio tiempo el polinomio minimal.

4. ¿Es la matriz A del Ejercicio 3 semejante, sobre el cuerpo de los números complejos, a una matriz diagonal?

5. Sea V un espacio de dimensión n y sea T un operador lineal sobre V. Supóngase que existe cierto entero positivo k tal que $T^k = 0$. Demostrar que $T^n = 0$.

6. Hallar una matriz 3×3 para la cual el polinomio minimal es x^2.

7. Sea n un entero positivo y sea V el espacio vectorial de los polinomios sobre R que tienen a lo más grado n (sin considerar el polinomio 0). Sea D el operador derivación sobre V. ¿Cuál es el polinomio minimal de D?

8. Sea P el operador en R^2 que proyecta cada vector sobre el eje x, paralelamente al eje y: $P(x, y) = (x, 0)$. Mostrar que P es lineal. ¿Cuál es el polinomio minimal de P?

9. Sea A una matriz $n \times n$, con polinomio característico

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}.$$

Demostrar que

$$c_1 d_1 + \cdots + c_k d_k = \text{traza}\,(A).$$

10. Sea V el espacio vectorial de las matrices $n \times n$ sobre el cuerpo F. Sea A una matriz $n \times n$ fija. Sea T el operador lineal sobre V definido por

$$T(B) = AB.$$

Demostrar que el polinomio minimal de T es el polinomio minimal de A.

11. Sean A y B matrices $n \times n$ sobre el cuerpo F. De acuerdo con el Ejercicio 9 de la Sección 6.1, las matrices AB y BA tienen los mismos valores propios. ¿Tienen también el mismo polinomio característico? ¿Tienen también el mismo polinomio minimal?

6.4. Subespacios invariantes

En esta sección introduciremos algunos conceptos útiles en el estudio de un operador lineal. Usaremos estas ideas para obtener descripciones de los operadores diagonalizables (y triangulables) en términos de sus polinomios minimales.

Definición. *Sea V un espacio vectorial y T un operador lineal sobre V. Si W es un subespacio de V, se dice que W es* **invariante por** *T si para todo vector α de W el vector $T\alpha$ está en W, es decir, si $T(W)$ está contenido en W.*

Ejemplo 6. Si T es cualquier operador lineal sobre V, entonces V es invariante por T, como también lo es el subespacio nulo. La imagen de T y el espacio nulo de T son también invariantes por T.

Ejemplo 7. Sea F un cuerpo y sea D el operador derivación sobre el espacio $F[x]$ de los polinomios sobre F. Sea n un entero positivo y sea W el subespacio de los polinomios de grado no mayor que n. Entonces W es invariante por D. Esto no es más que otra manera de decir que D es «decreciente en grado».

Ejemplo 8. He aquí una útil generalización del Ejemplo 6. Sea T un operador lineal sobre V. Sea U cualquier operador lineal sobre V que conmute con T; es decir, $TU = UT$. Sea W la imagen de U y sea N el espacio nulo de U. Ambos W y N son invariantes por T. Si α está en la imagen de U, por ejemplo, $\alpha = U\beta$, entonces $T\alpha = T(U\beta) = U(T\beta)$, de modo que $T\alpha$ está en la imagen de U. Si α está en N, entonces $U(T\alpha) = T(U\alpha) = T(0) = 0$; luego $T\alpha$ está en N.

Un tipo especial de operador que conmuta con T es un operador $U = g(T)$, donde g es un polinomio. Por ejemplo, podría tenerse $U = T - cI$, donde c es un valor propio de T. El espacio nulo de U ya nos es familiar. Se ve que este ejemplo incluye el hecho (obvio) de que el espacio de vectores propios de T, asociados al valor propio c, es invariante por T.

Ejemplo 9. Sea T el operador lineal sobre R^2 representado en la base ordenada canónica por la matriz

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Entonces los únicos subespacios de R^2 que son invariantes por T son R^2, y el subespacio nulo. Cualquier otro subespacio invariante tendrá necesariamente dimensión 1. Pero si W es el subespacio generado por algún vector no nulo α, que W es invariante por T quiere decir que α es un vector propio, pero A no tiene valores propios reales.

Cuando el subespacio W es invariante por el operador T, entonces T induce un operador lineal T_W en el espacio W. El operador lineal T_W está definido por $T_W(\alpha) = T(\alpha)$, para α en W; pero T_W es un objeto diferente de T, ya que su dominio es W y no V.

Cuando V es de dimensión finita, la invariancia de W por T tiene una interpretación matricial simple que acaso deba mencionarse aquí. Supóngase que se elige una base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ para V tal que $\mathcal{B}' = \{\alpha_1, \ldots, \alpha_n\}$ es una base ordenada de $W (r = \dim W)$. Sea $A = [T]_{\mathcal{B}}$ de modo que

$$T\alpha_j = \sum_{i=1}^{n} A_{ij}\alpha_i.$$

Como W es invariante por T, el vector $T\alpha_j$ pertenece a W para $j \leq r$. Esto quiere decir que

(6-9) $$T\alpha_j = \sum_{i=1}^{r} A_{ij}\alpha_i, \qquad j \leq r.$$

En otras palabras, $A_{ij} = 0$ si $j \leq r$ e $i > r$.

Esquemáticamente, A tiene la forma bloque

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \tag{6-10}$$

donde B es una matriz $r \times r$, C es una matriz $r \times (n - r)$ y D es una matriz $(n - r) \times (n - r)$. El lector deberá observar que, conforme a (6-9), la matriz B es precisamente la matriz del operador inducido T_W en la base ordenada $\mathcal{B}'$.

Muy a menudo se razonará respecto a T y T_W sin hacer uso de la forma bloque de la matriz A en (6-10). Pero deberá observarse cómo surgen ciertas relaciones entre T_W y T de esa forma bloque.

Lema. *Sea W un subespacio invariante para T. El polinomio característico para el operador restricción T_W divide el polinomio característico de T. El polinomio minimal de T_W divide al polinomio minimal de T.*

Demostración. Se tiene

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

donde $A = [T]_{\mathcal{B}}$ y $B = [T_W]_{\mathcal{B}'}$. Por la forma bloque de la matriz

$$\det (xI - A) = \det (xI - B) \det (xI - D).$$

Esto demuestra la afirmación respecto a los polinomios característicos. Obsérvese que se usó indistintamente I para representar matrices unidad en tres tamaños diferentes.

La k-ésima potencia de la matriz A tiene la forma bloque

$$A^k = \begin{bmatrix} B^k & C_k \\ 0 & D^k \end{bmatrix}$$

donde C_k es alguna matriz $r \times (n - r)$. Por tanto, cualquier polinomio que anule a A también anula a B (y también a D). Así, el polinomio minimal de B divide al polinomio minimal de A. ▮

Ejemplo 10. Sea T cualquier operador lineal en un espacio V de dimensión finita. Sea W el subespacio generado por *todos* los vectores propios de T. Sean $c_1, \ldots, c_k$ los valores propios distintos de T. Para cada i, sea W_i el espacio de los vectores característicos asociados con el valor propio c_i, y será $\mathcal{B}_i$ una base ordenada para W_i. El lema anterior al Teorema 2 dice que $\mathcal{B}' = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ es una base ordenada de W. En particular,

$$\dim W = \dim W_1 + \cdots + \dim W_k.$$

Sea $\mathcal{B}' = \{\alpha_1, \ldots, \alpha_r\}$ de modo que los primeros α forman la base $\mathcal{B}_1$, los siguientes, $\mathcal{B}_2$, y así sucesivamente. Entonces

$$T\alpha_i = t_i\alpha_i, \qquad i = 1, \ldots, r$$

donde $(t_1, \ldots, t_r) = (c_1, c_1, \ldots, c_1, \ldots, c_k, c_k, \ldots, c_k)$ con c_i repetidos dim W_i veces.

Ahora W es invariante por T, ya que para cada α en W se tiene

$$\alpha = x_1\alpha_1 + \cdots + x_r\alpha_r$$
$$T\alpha = t_1x_1\alpha_1 + \cdots + t_rx_r\alpha_r.$$

Eligiendo otros vectores $\alpha_{r+1}, \ldots, \alpha_n$ en V tales que $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ sea una base para V. La matriz de T, respecto de $\mathcal{B}$, tiene la forma bloque (6-10), y la matriz del operador restricción T_W, respecto de la base $\mathcal{B}'$ es

$$B = \begin{bmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & t_r \end{bmatrix}.$$

El polinomio característico de B (es decir, de T_W) es

$$g = (x - c_1)^{e_1} \cdots (x - c_k)^{e_k}$$

donde $e_i = \dim W_i$. Además g divide a f, el polinomio característico de T. Por tanto, la multiplicidad de c_i, como raíz de f, es a lo menos dim W_i.

Todo esto aclara el Teorema 2; dice solamente que T es diagonalizable si, y solo si, $r = n$, si, y solo si, $e_1 + \cdots + e_k = n$. Esto no ayuda mucho para el caso no diagonalizable, pues no se conocen las matrices C y D de (6-10).

Definición. *Sea W un subespacio invariante para T, y sea α un vector de V. El* T-**conductor de** α **en** W *es el conjunto $S_T(\alpha; W)$, que consta de todos los polinomios g (sobre el cuerpo escalar) tales que $g(T)\alpha$ está en W.*

Como el operador T se considera fijo a lo largo de la mayor parte de los análisis, se suprimirá normalmente el subíndice T y se escribirá $S(\alpha; W)$. Los autores llaman corrientemente a esta colección de polinomios el «relleno» (*stuffer*, en inglés; o *das einstopfende Ideal*, en alemán). «Conductor» es el término más corriente, preferido por aquellos que piensan en un operador $g(T)$ menos agresivo, que suavemente dirige al vector α en W. En el caso especial de $W = \{0\}$, el conductor se llama T-**anulador de** α.

Lema. *Si W es un subespacio invariante para T, entonces W es invariante por todo polinomio de T. Así, pues, para todo α de V, el conductor $S(\alpha; W)$ es un ideal en el álgebra de los polinomios $F[x]$.*

Demostración. Si β está en W, entonces $T\beta$ está en W. En consecuencia, $T(T\beta) = T^2\beta$ está en W. Por inducción, $T^k\beta$ está en W para cada k. Formando combinaciones lineales se ve que $f(T)\beta$ está en W para todo polinomio f.

La definición de $S(\alpha; W)$ toma sentido si W es un subconjunto de V. Si W es un subespacio, entonces $S(\alpha; W)$ es un subespacio de $F[x]$, ya que

$$(cf + g)(T) = cf(T) + g(T).$$

Si W es también invariante por T, sea g un polinomio en $S(\alpha; W)$; es decir, sea $g(T)\alpha$ en W. Si f es cualquier polinomio entonces $f(T)[g(T)\alpha]$ estará en W. Como

$$(fg)(T) = f(T)g(T)$$

fg está en $S(\alpha; W)$. Con lo que el conductor absorbe la multiplicación por cualquier polinomio. ∎

El generador mónico único del ideal $S(\alpha; W)$ es también llamado **T-conductor de α en W** (el **T-anulador** en caso de que $W = \{0\}$). El T-conductor de α en W es el polinomio mónico g de menor grado tal que $g(T)\alpha$ está en W. Un polinomio f está en $S(\alpha; W)$ si, y solo si, g divide a f. Nótese que el conductor $S(\alpha; W)$ siempre contiene el polinomio minimal de T; luego cada *T-conductor divide al polinomio minimal de T.*

Como primera ilustración de cómo usar el conductor $S(\alpha; W)$, se caracterizarán operadores triangulables. El operador lineal T se llama **triangulable** si existe una base ordenada en la cual T está representado por una matriz triangular.

Lema. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F. Sea T un operador lineal sobre V tal que el polinomio minimal de T sea un producto de factores lineales*

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}, \qquad c_i \text{ en } F.$$

Sea W un subespacio propio de V ($W \neq V$) invariante por T. Existe un vector α tal que

(a) *α no pertenece a W.*
(b) *$(T - cI)\alpha$ está en W, para algún valor propio c del operador T.*

Demostración. Lo que dice (a) y (b) es que el T-conductor de α en W es un polinomio lineal. Sea β cualquier vector en V que no está en W. Sea g el T-conductor de β en W. Entonces g divide a p, el polinomio minimal de T. Como β no está en W, el polinomio g no es constante. Luego

$$g = (x - c_1)^{e_1} \cdots (x - c_k)^{e_k}$$

donde al menos uno de los enteros e_i es positivo. Se elige j de modo que $e_j > 0$. Entonces $(x - c_j)$ divide a g:

$$g = (x - c_j)h.$$

Por la definición de g, el vector $\alpha = h(T)\beta$ no puede estar en W. Pero

$$\begin{aligned}(T - c_jI)\alpha &= (T - c_jI)h(T)\beta \\ &= g(T)\beta\end{aligned}$$

pertenece a W. ∎

Teorema 5. *Sea V un espacio de dimensión finita sobre el cuerpo F y sea T un operador lineal sobre V. Entonces T es triangulable si, y solo si, el polinomio minimal de T es producto de polinomios lineales sobre F.*

Demostración. Supóngase que el polinomio minimal está factorizado en la forma:

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}.$$

Por aplicación repetida del lema anterior se llega a una base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ en la cual la matriz que representa T es triangular superior

$$(6\text{-}11) \qquad [T]_{\mathcal{B}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}.$$

Ahora (6-11) solo dice que

$$(6\text{-}12) \qquad T\alpha_j = a_{1j}\alpha_1 + \cdots + a_{jj}\alpha_j, \qquad 1 \leq j \leq n$$

esto es, $T\alpha_j$ está en el subespacio generado por $\alpha_1, \ldots, \alpha_j$. Para determinar los $\alpha_1, \ldots, \alpha_n$, se comienza aplicando el lema al subespacio $W = \{0\}$, para obtener el vector α_1. Entonces aplicando el lema a W_1, el espacio generado por α_1, se obtiene α_2. A continuación se aplica el lema a W_2, el espacio generado por α_1 y α_2. Se continúa de este modo. Un punto se merece comentario especial. Después que se hayan determinado los $\alpha_1, \ldots, \alpha_i$, son las relaciones del tipo triangular (6-12) para $j = 1. \ldots, i$ las que aseguran que el subespacio generado por $\alpha_1, \ldots, \alpha_i$ es invariante por T.

Si T es triangulable, es evidente que el polinomio característico de T tiene la forma

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}, \qquad c_i \text{ en } F.$$

Basta observar la matriz triangular (6-11). Los elementos de la diagonal $a_{11}, \ldots, a_{nn}$ son los valores propios, con c_i repetidos d_i veces. Pero si f puede ser así factorizado, también puede serlo el polinomio minimal, ya que divide a f. ▮

Corolario. *Sea F un cuerpo algebraicamente cerrado, v.g., el cuerpo de los números complejos. Cada matriz $n \times n$ sobre F es semejante, sobre F, a una matriz triangular.*

Teorema 6. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea T un operador lineal sobre V. Entonces T es diagonalizable si, y solo si, el polinomio minimal de T tiene la forma*

$$p = (x - c_1) \cdots (x - c_k)$$

donde los $c_1, \ldots, c_k$ son elementos distintos de F.

Demostración. Se ha observado anteriormente que, si T es diagonalizable, su polinomio minimal es producto de factores lineales distintos (véase el análisis anterior al Ejemplo 4). Para demostrar el recíproco, sea W el subespacio generado por todos los vectores propios de T, y supóngase que $W \neq V$. Por

el lema empleado en la demostración del Teorema 5, existen un vector α que no está en W y un valor propio c_j de T tal que el vector

$$\beta = (T - c_jI)\alpha$$

pertenece a W. Como β está en W

$$\beta = \beta_1 + \cdots + \beta_k$$

donde $T\beta_i = c_i\beta_i$, $1 \leq i \leq k$ y, por tanto, el vector

$$h(T)\beta = h(c_1)\beta_1 + \cdots + h(c_k)\beta_k$$

está en W, para cada polinomio h.

Ahora, $p = (x - c_j)q$, para algún polinomio q. También

$$q - q(c_j) = (x - c_j)h.$$

Se tiene

$$q(T)\alpha - q(c_j)\alpha = h(T)(T - c_jI)\alpha = h(T)\beta.$$

Pero $h(T)\beta$ está en W, y como

$$0 = p(T)\alpha = (T - c_jI)q(T)\alpha$$

el vector $q(T)\alpha$ está en W. Por tanto, $q(c_j)\alpha$ está en W. Como α no está en W, se tiene $q(c_j) = 0$. Lo cual contradice el hecho de que p tiene raíces distintas. ▌

Al final de la Sección 6.7 se dará una demostración diferente del Teorema 6. Además de ser un resultado elegante, el Teorema 6 es útil en cuanto al cálculo. Supóngase que se tiene un operador lineal T representado por la matriz A en cierta base ordenada, y se desea saber si T es diagonalizable. Se calcula el polinomio característico f. Si se puede factorizar f:

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

se tienen dos métodos diferentes para determinar si T es diagonalizable o no. Un método es ver si (para todo i) se pueden hallar d_i vectores propios independientes asociados al valor propio c_i. El otro método es comprobar si $(T - c_1I) \cdots (T - c_kI)$ es o no el operador cero.

El Teorema 5 da una demostración diferente para el teorema de Cayley-Hamilton. Ese teorema es fácil para una matriz triangular. De donde, por medio del Teorema 5, se tiene el resultado para cualquier matriz sobre un cuerpo algebraicamente cerrado. Todo cuerpo es un subcuerpo de un cuerpo algebraicamente cerrado. Si se conoce ese hecho se puede tener una demostración del teorema de Cayley-Hamilton para las matrices sobre cualquier cuerpo. Si se acepta, por último, el teorema fundamental del álgebra (el cuerpo de los números complejos es algebraicamente cerrado), entonces el Teorema 5 da una demostración del teorema de Cayley-Hamilton para las matrices complejas, y esa demostración es independiente de la dada anteriormente.

Ejercicios

1. Sea T el operador lineal sobre R^2 cuya matriz en la base ordenada canónica es

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix}.$$

(a) Demostrar que los únicos subespacios de R^2 invariantes por T son R^2 y el subespacio nulo.

(b) Si U es el operador lineal en C^2, cuya matriz en la base ordenada canónica es A, demostrar que U tiene un subespacio unidimensional invariante.

2. Sea W un subespacio invariante para T. Demostrar, sin referirse a las matrices, que el polinomio minimal para el operador restricción T_W divide al polinomio minimal de T.

3. Sea c un valor propio de T y sea W el espacio de los vectores propios asociados al valor propio c. ¿Cuál es el operador restricción T_W?

4. Sea

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 2 & -2 & 2 \\ 2 & -3 & 2 \end{bmatrix}.$$

¿Es A semejante, sobre el cuerpo de los números reales, a una matriz triangular? Si es así, hallar tal matriz triangular.

5. Cada matriz A tal que $A^2 = A$ es semejante a una matriz diagonal.

6. Sea T un operador lineal diagonalizable sobre un espacio vectorial V de dimensión n y sea W un subespacio invariante por T. Demostrar que el operador restricción T_W es diagonalizable.

7. Sea T un operador lineal sobre un espacio vectorial de dimensión finita sobre el cuerpo de los números complejos. Demostrar que T es diagonalizable si, y solo si, T es anulado por algún polinomio sobre C que tiene raíces distintas.

8. Sea T un operador lineal sobre V. Si todo subespacio de V es invariante por T, entonces T es un múltiplo escalar del operador identidad.

9. Sea T el operador integración indefinida

$$(Tf)(x) = \int_0^x f(t)\, dt$$

sobre el espacio de las funciones continuas en el intervalo $[0, 1]$. ¿Es el espacio de las funciones polinomios invariantes por T? ¿El espacio de las funciones diferenciables? ¿El espacio de las funciones que se anulan para $x = \frac{1}{2}$?

10. Sea A una matriz 3×3 de elementos reales. Demostrar que si A no es semejante sobre R a una matriz triangular, entonces A es semejante sobre C a una matriz diagonal.

11. ¿Es verdadero o falso que si una matriz triangular A es semejante a una matriz diagonal, entonces A es diagonal?

12. Sea T un operador lineal sobre un espacio vectorial de dimensión finita sobre un cuerpo F algebraicamente cerrado. Sea f un polinomio sobre F. Demostrar que c es un valor propio de $f(T)$ si, y solo si, $c = f(t)$, donde t es un valor propio de T.

13. Sea V el espacio de las matrices $n \times n$ sobre F. Sea A una matriz $n \times n$ dada, sobre F. Sean T y U los operadores lineales sobre V definidos por

$$T(B) = AB$$
$$U(B) = AB - BA.$$

(a) ¿Es verdadero o falso que si A es diagonalizable (sobre F), entonces T es diagonalizable?

(b) ¿Es verdadero o falso que si A es diagonalizable, entonces U es diagonalizable?

6.5. *Triangulación simultánea; Diagonalización simultánea*

Sea V un espacio de dimensión finita y sea $\mathfrak{F}$ una familia de operadores lineales sobre V. Ocurre preguntar si se pueden triangular o diagonalizar simultáneamente los operadores de $\mathfrak{F}$, es decir, cuándo se puede encontrar una base $\mathcal{B}$ tal que todas las matrices $[T]_{\mathcal{B}}$, con T en $\mathfrak{F}$, sean triangulares (o diagonales). En el caso de la diagonalización, es necesario que $\mathfrak{F}$ sea una familia de operadores que conmuten: $UT = TU$, para todo T, U en $\mathfrak{F}$. Esto se desprende del hecho de que todas las matrices diagonales conmutan. Obviamente es también necesario que cada operador de $\mathfrak{F}$ sea un operador diagonalizable. Para la triangulación simultánea cada operador en $\mathfrak{F}$ debe ser triangulable. No es necesario que $\mathfrak{F}$ sea una familia de operadores que conmuten; sin embargo, esa condición es suficiente para la triangulación simultánea (si cada T es triangulable individualmente). Estos resultados se desprenden, con ligeras variaciones, de las demostraciones de los Teoremas 5 y 6.

El subespacio W es **invariante por** (la familia de los operadores) $\mathfrak{F}$ si W es invariante por cualquier operador en $\mathfrak{F}$

Lema. *Sea $\mathfrak{F}$ una familia de operadores lineales triangulables de V que conmutan. Sea W un subespacio propio de V invariante por $\mathfrak{F}$. Existe un vector α en V tal que*

(a) *α no pertenece a W;*

(b) *para cada T en $\mathfrak{F}$, el vector $T\alpha$ está en el subespacio generado por α y W.*

Demostración. No se perderá generalidad si se supone que $\mathfrak{F}$ contiene solo un número finito de operadores, por la siguiente observación. Sea $\{T_1, \ldots, T_r\}$ un subconjunto linealmente independiente maximal de $\mathfrak{F}$, es decir, una base del subespacio generado por $\mathfrak{F}$. Si α es un vector tal que (b) es válida para todo T_i, entonces (b) se cumplirá para todo operador que sea combinación lineal de los $T_1, \ldots, T_r$.

Por el lema anterior al Teorema 5 (este lema es para un operador solo), se puede hallar un vector β_1 (no en W) y un escalar c_1 tal que $(T_1 - c_1I)\beta_1$ esté en W. Sea V_1 la colección de todos los vectores β en V tales que $(T_1 - c_1I)\beta$ esté en W. Entonces V_1 es un subespacio de V que es propiamente más extenso

que W. Además, V_1 es invariante por $\mathfrak{F}$, por tal razón si T conmuta con T_1, entonces

$$(T_1 - c_1I)(T\beta) = T(T_1 - c_1I)\beta.$$

Si β está en V_1, entonces $(T_1 - c_1I)\beta$ está en W. Como W es invariante por toda T de $\mathfrak{F}$, se tiene que $T(T_1 - c_1I)\beta$ está en W; es decir, $T\beta$ en V_1, para todo β en V_1 y todo T en $\mathfrak{F}$.

Ahora W es un subespacio propio de V_1. Sea U_2 el operador lineal en V_1 que se obtiene restringiendo T_2 al subespacio V_1. El polinomio minimal de U_2 divide al polinomio minimal de T_2. Por tanto, se puede aplicar el lema anterior al Teorema 5 a tal operador y al subespacio invariante W. Se obtiene un vector β_2 en V_1 (no en W) y un escalar c_2 tal que $(T_2 - c_2I)\beta_2$ está en W. Obsérvese que

(a) β_2 no está en W.
(b) $(T_1 - c_1I)\beta_2$ está en W.
(c) $(T_2 - c_2I)\beta_2$ está en W.

Sea V_2 el conjunto de todos los vectores β en V_1 tal que $(T_2 - c_2I)\beta$ está en W. Entonces V_2 es invariante por $\mathfrak{F}$. Aplíquese el lema anterior al Teorema 5 a U_3, restricción de T_3 a V_2. Si se sigue así se llega a obtener un vector $\alpha = \beta_r$ (no en W) tal que $(T_j - c_jI)\alpha$ está en W, $j = 1, \ldots, r$. ▮

Teorema 7. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F. Sea $\mathfrak{F}$ una familia conmutativa de operadores lineales triangulables sobre V. Existe una base ordenada de V tal que todo operador de $\mathfrak{F}$ está representado por una matriz triangular en esa base.*

Demostración. Dado el lema recién demostrado, este teorema tiene la misma demostración que la dada para el Teorema 5, si se remplaza T por $\mathfrak{F}$. ▮

Corolario. *Sea $\mathfrak{F}$ una familia conmutativa de matrices $n \times n$ sobre un cuerpo F algebraicamente cerrado. Existe una matriz P, no singular, $n \times n$, con elementos en F, tal que $P^{-1}AP$ es triangular superior, para cada matriz A en $\mathfrak{F}$.*

Teorema 8. *Sea $\mathfrak{F}$ una familia conmutativa de operadores lineales diagonalizables en el espacio vectorial V de dimensión finita. Existe una base ordenada de V tal que todo operador de $\mathfrak{F}$ está representado en esa base por una matriz diagonal.*

Demostración. Se puede demostrar este teorema usando el lema anterior al Teorema 7 para el caso diagonalizable, de igual modo como se adaptó el lema anterior al Teorema 5 para demostrar el caso diagonalizable para el Teorema 6. Sin embargo, a esta altura es más fácil proceder por inducción sobre la dimensión de V.

Si $\dim V = 1$, no hay nada que demostrar. Supóngase que el teorema es válido para espacios vectoriales de dimensión menor que n y sea V un espacio vectorial de dimensión n. Elíjase cualquier T en $\mathfrak{F}$ que no sea un múltiplo escalar

de la identidad. Sean $c_1, \ldots, c_k$ los valores propios distintos de T, y (para cada i) sea W_i el espacio nulo de $(T - c_i I)$. Se fija un índice i. Entonces W_i es invariante por cada operador que conmuta con T. Sea $\mathfrak{F}_i$ la familia de operadores lineales en W_i que se obtiene por la restricción de los operadores de $\mathfrak{F}$ al subespacio (invariante) W_i. Cada operador de $\mathfrak{F}_i$ es diagonalizable, porque su polinomio minimal divide al polinomio minimal del correspondiente operador en $\mathfrak{F}$. Como $\dim W_i < \dim V$, los operadores en $\mathfrak{F}_i$ pueden diagonalizarse simultáneamente. Esto es, W_i tiene una base $\mathcal{B}_i$ que consta de vectores que son simultáneamente vectores propios para cada operador en $\mathfrak{F}_i$.

Como T es diagonalizable, el lema anterior al Teorema 2 dice que $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ es una base para V. Esa es la base que se buscaba. ▌

Ejercicios

1. Hallar una matriz real inversible P tal que $P^{-1}AP$ y $P^{-1}BP$ sean ambas diagonales, donde A y B son las matrices reales

(a) $$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -8 \\ 0 & -1 \end{bmatrix}$$

(b) $$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}.$$

2. Sea $\mathfrak{F}$ una familia de matrices complejas 3×3 que conmutan. ¿Cuántas matrices linealmente independientes puede tener $\mathfrak{F}$? ¿Qué se puede decir del caso $n \times n$?

3. Sea T un operador lineal en un espacio de dimensión n y supóngase que T tiene n valores propios distintos. Demostrar que todo operador lineal que conmuta con T es un polinomio en T.

4. Sean A, B, C y D matrices complejas $n \times n$ que conmutan. Sea E la matriz $2n \times 2n$

$$E = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

Demostrar que $\det E = (AD - BC)$.

5. Sean F un cuerpo, n un entero positivo y V el espacio de las matrices $n \times n$ sobre F. Si A es una matriz $n \times n$ fija sobre F, sea T_A el operador lineal sobre V definido por $T_A(B) = AB - BA$. Considerar la familia de operadores lineales T_A que se obtiene haciendo variar A sobre todas las matrices diagonales. Demostrar que los operadores de esa familia son simultáneamente diagonalizables.

6.6. *Descomposiciones en suma directa*

Al continuar con el análisis de un operador lineal se deben formular las ideas de un modo algo más refinado – menos en términos de matrices y más en términos de subespacios. Cuando comenzamos este capítulo, describimos su propósito del siguiente modo: encontrar una base ordenada en la cual la matriz de T toma una forma especialmente simple. Ahora describiremos nues-

tro propósito como sigue: descomponer el espacio base V en una suma de subespacios invariantes para T tal que los operadores restricción a esos subespacios sean simples.

Definición. *Sean $W_1, \ldots, W_k$ subespacios de un espacio vectorial V. Se dice que $W_1, \ldots, W_k$ son* **independientes** *si*

$$\alpha_1 + \cdots + \alpha_k = 0, \qquad \alpha_i \text{ en } W_i$$

implica que cada α_i es 0.

Para $k = 2$, el sentido de la independencia es la intersección $\{0\}$; es decir, W_1 y W_2 son independientes si, y solo si, $W_1 \cap W_2 = \{0\}$. Si $k > 2$, la independencia de $W_1, \ldots, W_k$ dice mucho más que $W_1 \cap \cdots \cap W_k = \{0\}$. Dice que cada W_j encuentra la suma de los otros subespacios W_i solo en el vector nulo.

El significado de independencia es el siguiente. Sea $W_1 + \cdots + W_k$ el subespacio generado por $W_1, \ldots, W_k$. Cada vector α de W puede ser expresado como suma

$$\alpha = \alpha_1 + \cdots + \alpha_k, \qquad \alpha_i \text{ en } W_i.$$

Si $W_1, \ldots, W_k$ son independientes, entonces esa expresión de α es única; en efecto, si

$$\alpha = \beta_1 + \cdots + \beta_k, \qquad \beta_i \text{ en } W_i$$

entonces $0 = (\alpha_1 - \beta_1) + \cdots + (\alpha_k - \beta_k)$, luego $\alpha_i - \beta_i = 0$, $i = 1, \ldots, k$. Así, si $W_1, \ldots, W_k$ son independientes, se puede operar con los vectores de W como k-tuples $(\alpha_1, \ldots, \alpha_k)$, con α_i en W_i, en la misma forma como se opera con los vectores de R^k como k-tuples de números.

Lema. *Sea V un espacio vectorial de dimensión finita. Sean $W_1, \ldots, W_k$ subespacios de V y sea $W = W_1 + \cdots + W_k$. Las siguientes afirmaciones son equivalentes.*

(a) *$W_1, \ldots, W_k$ son independientes.*
(b) *Para todo j, $2 \le j \le k$, se tiene*

$$W_j \cap (W_1 + \cdots + W_{j-1}) = \{0\}.$$

(c) *Si $\mathcal{B}_i$ es una base ordenada de W_i, $1 \le i \le k$, entonces la sucesión $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ es una base ordenada para W.*

Demostración. Supóngase (a). Sea α un vector de la intersección $W_j \cap (W_1 + \cdots + W_{j-1})$. Entonces existen vectores $\alpha_1, \ldots, \alpha_{j-1}$ con α_i en W_i tales que $\alpha = \alpha_1 + \cdots + \alpha_{j-1}$. Como

$$\alpha_1 + \cdots + \alpha_{j-1} + (-\alpha) + 0 + \cdots + 0 = 0$$

y como $W_1, \ldots, W_k$ son independientes, debe tenerse que $\alpha_1 = \alpha_2 \cdots = \alpha_{j-1} = \alpha = 0$.

Ahora obsérvese que (b) implica a (a). Supóngase

$$0 = \alpha_1 + \cdots + \alpha_k, \qquad \alpha_i \text{ en } W_i.$$

Sea j el mayor entero i tal que $\alpha_i \neq 0$. Entonces

$$0 = \alpha_1 + \cdots + \alpha_j, \qquad \alpha_j \neq 0.$$

Así, $\alpha_j = -\alpha_1 - \cdots - \alpha_{j-1}$, es un vector no nulo en $W_j \cap (W_1 + \cdots + W_{j-1})$.

Ahora que sabemos que (a) y (b) son equivalentes, veamos por qué (a) y (c) son equivalentes. Supóngase (a). Sea $\mathcal{B}_i$ una base para W_i, $1 \leq i \leq k$, y sea $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$. Cualquier relación lineal entre los vectores de $\mathcal{B}$ tendrá la forma

$$\beta_1 + \cdots + \beta_k = 0$$

donde β_i es cierta combinación lineal de los vectores en $\mathcal{B}_i$. Como $W_1, \ldots, W_k$ son independientes, cada β_i es 0. Como cada $\mathcal{B}_i$ es independiente, la relación que tenemos entre los vectores en $\mathcal{B}$ es la relación trivial.

Se deja la demostración que (c) implica (a) para los ejercicios (Ejercicio 2). ▌

Si cualquiera (y, por tanto, todas) de las condiciones del último lema se cumple, se dice que la suma $W = W_1 + \cdots + W_k$ es **directa** o que W es la **suma directa** de $W_1, \ldots, W_k$, y se escribe

$$W = W_1 \oplus \cdots \oplus W_k.$$

En la literatura pertinente el lector puede encontrar esta suma directa como suma independiente o suma directa interna de $W_1, \ldots, W_k$.

Ejemplo 11. Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea $\{\alpha_1, \ldots, \alpha_n\}$ una base de V. Si W_i es el subespacio unidimensional generado por α_i, entonces $V = W_1 \oplus \cdots \oplus W_n$.

Ejemplo 12. Sea n un entero positivo y F un subcuerpo de los números complejos, y sea V el espacio de todas las matrices $n \times n$ sobre F. Sea W_1 el subespacio de todas las matrices **simétricas**, es decir, matrices A tales que $A^t = A$. Sea W_2 el subespacio de todas las matrices **antisimétricas**, es decir, matrices A tales que $A^t = -A$. Entonces $V = W_1 \oplus W_2$. Si A es cualquier matriz de V, la expresión única para A como suma de matrices, una en W_1 y la otra en W_2, es

$$\begin{aligned} A &= A_1 + A_2 \\ A_1 &= \tfrac{1}{2}(A + A^t) \\ A_2 &= \tfrac{1}{2}(A - A^t). \end{aligned}$$

Ejemplo 13. Sea T cualquier operador lineal sobre un espacio V de dimensión finita. Sean $c_1, \ldots, c_k$ los valores propios distintos de T, y sea W_i el espacio de los vectores propios asociados a los valores propios c_i. Entonces $W_1, \ldots, W_k$ son independientes. Véase el lema anterior al Teorema 2. En particular, si T es diagonalizable, entonces $V = W_1 \oplus \cdots \oplus W_k$.

Definición. *Si V es un espacio vectorial, una* **proyección** *de V es un operador lineal E sobre V tal que $E^2 = E$.*

Supóngase que E sea una proyección. Sea R la imagen de E y sea N el espacio nulo de E.

1. El vector β está en la imagen R si, y solo si, $E\beta = \beta$. Si $\beta = E\alpha$, entonces $E\beta = E^2\alpha = E\alpha = \beta$. Recíprocamente, si $\beta = E\beta$, entonces (naturalmente) β está en la imagen de E.
2. $V = R \oplus N$.
3. La expresión única de α, como suma de vectores en R y en N, es $\alpha = E\alpha + (\alpha - E\alpha)$.

De (1), (2) y (3) es fácil ver lo siguiente. Si R y N son subespacios de V tales que $V = R \oplus N$, existe un, y solo un, operador proyección E que tiene por imagen R y por espacio nulo N. Este operador se llama **proyección sobre R según o paralelamente a** N.

Cualquier proyección E es (trivialmente) diagonalizable. Si $\{\alpha_1, \ldots, \alpha_r\}$ es una base de R y $\{\alpha_{r+1}, \ldots, \alpha_n\}$ es una base de N, entonces la base $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ diagonaliza a E:

$$[E]_{\mathcal{B}} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

donde I es la matriz unidad $r \times r$. Esto ayudará a explicar parte de la terminología relacionada con las proyecciones. El lector deberá ver las diferentes situaciones en el plano R^2 (o el espacio tridimensional, R^3), para convencerse de que la proyección en R, según N, aplica todo vector en R por proyección paralela a N.

Las proyecciones pueden ser empleadas para describir las descomposiciones en suma directa del espacio V. En efecto, supóngase que $V = W_1 \oplus \cdots \oplus W_k$. Para todo j se define un operador E_j sobre V. Sea α en V, v. g., $\alpha = \alpha_1 + \cdots + \alpha_k$, con α_i en W_i. Defínase $E_j\alpha = \alpha_j$. Entonces E_j es una ley bien definida. Es fácil ver que E_j es lineal, que la imagen de E_j es W_j y que $E_j^2 = E_j$. El espacio nulo de E_j es el subespacio

$$(W_1 + \cdots + W_{j-1} + W_{j+1} + \cdots + W_k)$$

en efecto, la afirmación de que $E_j\alpha = 0$ solo dice que $\alpha_j = 0$; es decir, que α es efectivamente una suma directa de vectores de los espacios W_i, con $i \neq j$. En términos de las proyecciones E_j, se tiene

(6-13) $$\alpha = E_1\alpha + \cdots + E_k\alpha$$

para todo α de V. Lo que (6-13) dice es que

$$I = E_1 + \cdots + E_k.$$

Obsérvese también que si $i \neq j$, entonces $E_iE_j = 0$, ya que la imagen de E_j es el subespacio W_j, contenido en el espacio nulo de E_i. Se resume ahora lo dicho y se enuncia y se demuestra un recíproco.

Teorema 9. *Si* $V = W_1 \oplus \cdots \oplus W_k$, *entonces existen k operadores lineales* $E_1, \ldots, E_k$ *sobre* V *tales que*

(i) *todo* E_i *es una proyección* $(E_i^2 = E_i)$;
(ii) $E_iE_j = 0$ *si* $i \neq j$;
(iii) $I = E_1 + \cdots + E_k$;
(iv) *la imagen de* E_i *es* W_i.

Recíprocamente, si $E_1, \ldots, E_k$ *son los operadores lineales que satisfacen las condiciones* (i), (ii) *y* (iii) *y si se hace que* W_i *sea la imagen de* E_i, *entonces* $V = W_1 \oplus \cdots \oplus W_k$.

Demostración. Solo se necesita probar el recíproco. Supóngase que $E_1, \ldots, E_k$ son operadores lineales sobre V que satisfacen las tres primeras condiciones y sea W_i la imagen de E_i. Entonces, claramente

$$V = W_1 + \cdots + W_k;$$

en efecto, por la condición (iii) se tiene

$$\alpha = E_1\alpha + \cdots + E_k\alpha$$

para todo α de V y $E_i\alpha$ está en W_i. Esta expresión de α es única, pues si

$$\alpha = \alpha_1 + \cdots + \alpha_k$$

con α_i en W_i, v.g., $\alpha_i = E_i\beta_i$, entonces, por (i) y (ii) se tiene

$$\begin{aligned} E_j\alpha &= \sum_{i=1}^{k} E_j\alpha_i \\ &= \sum_{i=1}^{k} E_jE_i\beta_i \\ &= E_j^2\beta_j \\ &= E_j\beta_j \\ &= \alpha_j. \end{aligned}$$

Esto muestra que V es la suma directa de los W_i. ∎

Ejercicios

1. Sea V un espacio vectorial de dimensión finita y sea W_1 un subespacio de V. Demostrar que existe un subespacio W_2 de V tal que $V = W_1 \oplus W_2$.

2. Sea V un espacio vectorial de dimensión finita y sean $W_1, \ldots, W_k$ subespacios de V tales que

$$V = W_1 + \cdots + W_k \quad \text{y} \quad \dim V = \dim W_1 + \cdots + \dim W_k.$$

Demostrar que $V = W_1 \oplus \cdots \oplus W_k$.

3. Hallar una proyección E que proyecte R^2 sobre el subespacio generado por $(1, -1)$ según el subespacio generado por $(1, 2)$.

4. Si E_1 y E_2 son proyecciones sobre subespacios independientes, entonces $E_1 + E_2$ es una proyección. ¿Es verdadero o falso?

5. Si E es una proyección y f un polinomio, entonces $f(E) = aI + bE$. ¿Qué son a y b en términos de los coeficientes de f?

6. Si un operador diagonalizable tiene solo los valores propios 0 y 1 es una proyección ¿Es o no cierto?

7. Demostrar que si E es la proyección de R según N, entonces $(I - E)$ es la proyección en N según R.

8. Sean $E_1, \ldots, E_k$ operadores lineales sobre el espacio V de modo que $E_1 + \cdots + E_k = I$.
(a) Demostrar que si $E_iE_j = 0$, para $i \neq j$, entonces $E_i^2 = E_i$, para cada i.
(b) En el caso $k = 2$, demostrar el recíproco de (a). Esto es, si $E_1 + E_2 = I$, además $E_1^2 = E_1$ y $E_2^2 = E_2$, entonces $E_1E_2 = 0$.

9. Sean V un espacio vectorial real y E un operador lineal idempotente en V, es decir, una proyección. Demostrar que $(I + E)$ es inversible. Hallar $(I + E)^{-1}$.

10. Sea F un subcuerpo de los números complejos (o un cuerpo de característica cero). Sea V un espacio vectorial de dimensión finita sobre F. Supóngase que $E_1, \ldots, E_k$ son proyecciones de V y que $E_1 + \cdots + E_k = I$. Demostrar que $E_iE_j = 0$, para $i \neq j$. (*Sugerencia*: Usar la función traza y preguntarse qué es la traza de una proyección.)

11. Sea V un espacio vectorial, sean $W_1, \ldots, W_k$ subespacios de V y considérese que

$$V_j = W_1 + \cdots + W_{j-1} + W_{j+1} + \cdots + W_k.$$

Supóngase que $V = W_1 \oplus \cdots \oplus W_k$. Demostrar que el espacio dual V^* tiene la descomposición en suma directa $V^* = V_1^0 \oplus \cdots \oplus V_k^0$.

6.7. *Sumas directas invariantes*

Estamos interesados especialmente en descomposiciones en suma directa $V = W_1 \oplus \cdots \oplus W_k$, donde cada uno de los subespacios W_i es invariante bajo un operador lineal T dado. Dada tal descomposición de V, T induce un operador lineal T_i sobre cada W_i por restricción. El efecto de T es entonces el siguiente. Si α es un vector de V, se tienen vectores únicos $\alpha_1, \ldots, \alpha_k$, con α_i en W_i, de modo que

$$\alpha = \alpha_1 + \cdots + \alpha_k$$

y entonces

$$T\alpha = T_1\alpha_1 + \cdots + T_k\alpha_k.$$

Esta situación se describirá diciendo que T es la **suma directa** de los operadores $T_1, \ldots, T_k$. Al usar esta terminología, debe recordarse que los T_i no son operadores lineales sobre el espacio V, pero sí sobre los distintos subespacios W_i. El hecho de que $V = W_1 \oplus \cdots \oplus W_k$ permite asociar a cada α en V un k-tuple único $(\alpha_1, \ldots, \alpha_k)$ de vectores α_i en W_i (por $\alpha = \alpha_1 + \cdots + \alpha_k$) de modo que

podamos efectuar las operaciones lineales en V trabajando en los subespacios individuales W_i. El hecho de que cada W_i sea invariante por T permite considerar el efecte de T como efecto independiente de los operadores T_i sobre los subespacios W_i. Nuestro propósito ahora es estudiar T, encontrando la descomposición en suma directa invariante en que los T_i son operadores de una naturaleza elemental.

Antes de ver un ejemplo, observemos el análogo matricial de esta situación. Supóngase que seleccionamos una base ordenada $\mathcal{B}_i$ para cada W_i y sea $\mathcal{B}$ la base ordenada de V constituida por la unión de las $\mathcal{B}_i$ dispuestas en el orden $\mathcal{B}_1, \ldots, \mathcal{B}_k$, de modo que $\mathcal{B}$ sea una base para V. Del estudio referente al análogo matricial para un subespacio invariante simple, es fácil ver que si $A = [T]_{\mathcal{B}}$ y $A_i = [T_i]_{\mathcal{B}_i}$, entonces A tiene la forma bloque

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix}. \tag{6-14}$$

En (6-14) A_i es una matriz $d_i \times d_i$ ($d_i = \dim W_i$), y los 0 son símbolos para los bloques rectangulares con escalares 0 de varias dimensiones. Es también apropiado describir (6-14) diciendo que A es la **suma directa** de las matrices $A_i, \ldots, A_k$.

Más a menudo describiremos el subespacio W_i por medio de las proyecciones asociadas E_i (Teorema 9). Por tanto, se necesita expresar la invarianza del subespacio W_i en términos de los E_i.

Teorema 10. *Sea T un operador lineal sobre el espacio V y sean $W_1, \ldots, W_k$ y $E_1, \ldots, E_k$ como en el Teorema 9. Entonces una condición necesaria y suficiente para que cada subéspacio W_i sea invariante por T es que T conmute con cada una de las proyecciones E_i; es decir,*

$$TE_i = E_iT, \qquad i = 1, \ldots, k.$$

Demostración. Supóngase que T conmute con cada E_i. Sea α en W_j. Entonces $E_j\alpha = \alpha$, y

$$\begin{aligned} T\alpha &= T(E_j\alpha) \\ &= E_j(T\alpha) \end{aligned}$$

que muestra que $T\alpha$ está en la imagen de E_j, es decir, que W_j es invariante por T.

Supóngase ahora que todo W_i es invariante por T. Se ha de demostrar que $TE_j = E_jT$. Sea α cualquier vector de V. Entonces

$$\begin{aligned} \alpha &= E_1\alpha + \cdots + E_k\alpha \\ T\alpha &= TE_1\alpha + \cdots + TE_k\alpha. \end{aligned}$$

Como $E_i\alpha$ está en W_i, que es invariante por T, se debe tener que $T(E_i\alpha) = E_i\beta_i$ para cierto vector β_i. Entonces

$$\begin{aligned} E_jTE_i\alpha &= E_jE_i\beta_i \\ &= \begin{cases} 0, & \text{si } i \neq j \\ E_j\beta_j, & \text{si } i = j. \end{cases} \end{aligned}$$

Con lo que

$$\begin{aligned} E_jT\alpha &= E_jTE_1\alpha + \cdots + E_jTE_k\alpha \\ &= E_j\beta_j \\ &= TE_j\alpha. \end{aligned}$$

Esto vale para cada α de V, o sea $E_jT = TE_j$. ▮

Describiremos ahora un operador diagonalizable T en el lenguaje de las descomposiciones en suma directa invariante (proyecciones que conmutan con T). Esto será de gran utilidad para entender después unos teoremas de descomposición más profundos. El lector puede pensar que la descripción que se ha dado es muy complicada, en comparación con la formulación matricial o la simple afirmación de que los vectores propios de T generan el espacio total. Pero se debe tener presente que esta es la primera experiencia con un método muy efectivo por medio del cual varios problemas concernientes a subespacios, bases, matrices y entes similares pueden reducirse a cálculos algebraicos con operadores lineales. Con un poco de experiencia, la eficacia y elegancia de este método de razonamiento llegarán a ser claras.

Teorema 11. *Sea T un operador lineal sobre un espacio de dimensión finita V. Si T es diagonalizable y si $c_1, \ldots, c_k$ son los valores propios distintos de T, entonces existen operadores lineales $E_1, \ldots, E_k$ en V tales que*

(i) $T = c_1E_1 + \cdots + c_kE_k$;
(ii) $I = E_1 + \cdots + E_k$;
(iii) $E_iE_j = 0$, $i \neq j$;
(iv) $E_i^2 = E_i$ (*E_i es una proyección*);
(v) *la imagen de E_i es el espacio propio de T, asociado a c_i.*

Recíprocamente, si existen k escalares distintos $c_1, \ldots, c_k$ y k operadores lineales no nulos $E_1, \ldots, E_k$ que satisfacen (i), (ii) *y* (iii), *entonces T es diagonalizable, $c_1, \ldots, c_k$ son los valores propios distintos de T y las condiciones* (iv) *y* (v) *también se cumplen.*

Demostración. Supóngase que T es diagonalizable, con valores propios distintos $c_1, \ldots, c_k$. Sea W_i el subespacio de los vectores propios asociados al valor propio c_i. Como se sabe,

$$V = W_1 \oplus \cdots \oplus W_k.$$

Sean $E_1, \ldots, E_k$ las proyecciones asociadas con esta descomposición, como en el Teorema 9. Entonces (ii), (iii), (iv) y (v) se cumplen. Para verificar (i) se procede como sigue. Para cada α en V,

$$\alpha = E_1\alpha + \cdots + E_k\alpha$$

y así,

$$\begin{aligned} T\alpha &= TE_1\alpha + \cdots + TE_k\alpha \\ &= c_1E_1\alpha + \cdots + c_kE_k\alpha. \end{aligned}$$

En otras palabras, $T = c_1E_1 + \cdots + c_kE_k$.

Ahora supóngase que tenemos un operador lineal T junto con escalares c_i distintos y operadores no nulos E_i que satisfacen las condiciones (i), (ii) y (iii). Como $E_iE_j = 0$, cuando $i \neq j$, multiplicando ambos miembros de $I = E_1 + \cdots + E_k$ por E_i, se tiene de inmediato que $E_i^2 = E_i$. Multiplicando $T = c_1E_1 + \cdots + c_kE_k$ por E_i se tiene entonces que $TE_i = c_iE_i$, lo que muestra que cada vector de la imagen de E_i está en el espacio nulo de $(T - c_iI)$. Como se ha supuesto que $E_i \neq 0$, esto demuestra que existe un vector no nulo en el espacio nulo de $(T - c_iI)$, es decir, que c_i es un valor propio de T. Además, los c_i son todos los valores propios de T; en efecto, si c es otro escalar cualquiera, entonces

$$T - cI = (c_1 - c)E_1 + \cdots + (c_k - c)E_k$$

con lo que si $(T - cI)\alpha = 0$, se debe tener que $(c_i - c)E_i\alpha = 0$. Si α no es el vector nulo, entonces $E_i\alpha \neq 0$, para algún i, con lo que para ese i se tiene que $c_i - c = 0$.

Ciertamente T es diagonalizable, ya que se ha visto que todo vector no nulo, de la imagen de E_i, es un vector propio de T, y el que $I = E_1 + \cdots + E_k$ muestra que estos vectores propios generan V. Todo lo que queda por demostrar es que el espacio nulo de $(T - c_iT)$ es exactamente la imagen de E_i. Pero esto es evidente, porque si $T\alpha = c_i\alpha$ entonces

$$\sum_{j=1}^{k} (c_j - c_i)E_j\alpha = 0$$

luego

$$(c_j - c_i)E_j\alpha = 0 \qquad \text{para cada } j$$

y entonces

$$E_j\alpha = 0, \qquad j \neq i.$$

Como $\alpha = E_1\alpha + \cdots + E_k\alpha$ y $E_j\alpha = 0$, para $j \neq i$, se tiene que $\alpha = E_i\alpha$, lo que demuestra que α está en la imagen de E_i. ∎

Una parte del Teorema 9 dice que para un operador diagonalizable T los escalares $c_1, \ldots, c_k$ y los operadores $E_1, \ldots, E_k$ están unívocamente determinados por las condiciones (i), (ii), (iii), por ser los c_i distintos y porque los E_i son no nulos. Una de las ventajas más notables de la descomposición $T = c_1E_1 + \cdots + c_kE_k$ es que si g es cualquier polinomio sobre el cuerpo F, entonces

$$g(T) = g(c_1)E_1 + \cdots + g(c_k)E_k.$$

Dejamos los detalles de la demostración al lector. Para ver cómo se demuestra debe calcularse T^r para cada entero positivo r. Por ejemplo,

$$\begin{aligned} T^2 &= \sum_{i=1}^{k} c_iE_i \sum_{j=1}^{k} c_jE_j \\ &= \sum_{i=1}^{k} \sum_{j=1}^{k} c_ic_jE_iE_j \\ &= \sum_{i=1}^{k} c_i^2E_i^2 \\ &= \sum_{i=1}^{k} c_i^2E_i. \end{aligned}$$

El lector debe comparar esto con $g(A)$, donde A es una matriz diagonal; en ese caso $g(A)$ es simplemente la matriz diagonal cuyos elementos de la diagonal son $g(A_{11}), \ldots, g(A_{nn})$.

Querríamos en particular hacer notar qué pasa cuando se aplican los polinomios de Lagrange correspondientes a los escalares $c_1, \ldots, c_k$:

$$p_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}.$$

Se tiene $p_j(c_i) = \delta_{ij}$, lo que quiere decir que

$$\begin{aligned} p_j(T) &= \sum_{i=1}^{k} \delta_{ij} E_i \\ &= E_j. \end{aligned}$$

Así las proyecciones E_j no solo conmutan con T, sino que también son polinomios en T.

Tales cálculos con los polinomios en T pueden ser usados para dar otra demostración del Teorema 6 que caracteriza los operadores diagonalizables en términos de sus polinomios minimales. La demostración es enteramente independiente de la anterior.

Si T es diagonalizable, $T = c_1E_1 + \cdots + c_kE_k$, entonces

$$g(T) = g(c_1)E_1 + \cdots + g(c_k)E_k$$

para todo polinomio g. Así $g(T) = 0$ si, y solo si, $g(c_i) = 0$, para todo i. En particular, el polinomio minimal de T es

$$p = (x - c_1) \cdots (x - c_k).$$

Ahora supóngase que T sea un operador lineal con polinomio minimal $p = (x - c_1) \cdots (x - c_k)$, donde $c_1, \ldots, c_k$ son elementos distintos del cuerpo escalar. Se forman los polinomios de Lagrange

$$p_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}.$$

Recordemos del Capítulo 4 que $p_j(c_i) = \delta_{ij}$ y para cualquier polinomio g de grado menor o igual a $(k - 1)$ tenemos

$$g = g(c_1)p_1 + \cdots + g(c_k)p_k.$$

Tomando g como el polinomio escalar 1 y luego el polinomio x, tenemos

$$\begin{aligned} 1 &= p_1 + \cdots + p_k \\ x &= c_1p_1 + \cdots + c_kp_k. \end{aligned} \tag{6-15}$$

(El lector advertido notará que la aplicación a x puede no ser válida, pues k puede ser 1. Pero si $k = 1$, T es un múltiplo escalar de la identidad, por tanto, diagonalizable.) Ahora sea $E_j = p_j(T)$. De (6-15) tenemos

$$\begin{aligned} I &= E_1 + \cdots + E_k \\ T &= c_1E_1 + \cdots + c_kE_k. \end{aligned} \tag{6-16}$$

Obsérvese que si $i \neq j$, entonces $p_i p_j$ es divisible por el polinomio minimal p, pues $p_i p_j$ contiene a cada $(x - c_r)$ como factor. Así,

$$E_i E_j = 0, \qquad i \neq j. \tag{6-17}$$

Debemos notar un aspecto más, a saber, que $E_i \neq 0$, para todo i. Esto es porque p es el polinomio minimal de T y así no podemos tener que $p_i(T) = 0$, ya que p_i es de grado menor que el grado de p. Este último comentario, junto con (6-16), (6-17) y el hecho de que los c_i son distintos, permite aplicar el Teorema 11 para concluir que T es diagonalizable. ▌

Ejercicios

1. Sea E una proyección de V y sea T un operador lineal sobre V. Demostrar que la imagen de E es invariante por T si, y solo si, $ETE = TE$. Demostrar que ambos, la imagen y el espacio nulo de E, son invariantes por T si, y solo si, $ET = TE$.

2. Sea T el operador lineal sobre R^2, cuya matriz en la base ordenada canónica es

$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}.$$

Sea W_1 el subespacio de R^2 generado por el vector $\epsilon_1 = (1, 0)$.

(a) Demostrar que W_1 es invariante por T.

(b) Demostrar que no existe un subespacio W_2 que es invariante por T y que es complementario de W_1:

$$R^2 = W_1 \oplus W_2.$$

(Comparar con el Ejercicio 1 de la Sección 6.5).

3. Sea T un operador lineal sobre un espacio vectorial de dimensión finita. Sea R la imagen de T y sea N el espacio nulo de T. Demostrar que R y N son independientes si, y solo si, $V = R \oplus N$.

4. Sea T un operador lineal sobre V. Supóngase que $V = W_1 \oplus \cdots \oplus W_k$, donde cada W_i es invariante por T. Sea T_i el operador inducido (restricción) sobre W_i.

(a) Demostrar que $\det(T) = \det(T_1) \cdots \det(T_k)$.

(b) Demostrar que el polinomio característico de T es producto de los polinomios característicos de $T_1, \ldots, T_k$.

(c) Demostrar que el polinomio minimal de T es el mínimo común múltiplo de los polinomios minimales de $T_1, \ldots, T_k$. (*Sugerencia:* Demostrar y emplear el hecho correspondiente para sumas directas de matrices.)

5. Sea T el operador lineal diagonalizable sobre R^2 que se examinó en el Ejemplo 3 de la Sección 6.2. Usar los polinomios de Lagrange para escribir la matriz representante A en la forma $A = E_1 + 2E_2$, $E_1 + E_2 = I$, $E_1E_2 = 0$.

6. Sea A la matriz 4×4 del Ejemplo 6 de la sección 6.3. Hallar las matrices E_1, E_2, E_3 de modo que $A = c_1E_1 + c_2E_2 + c_3E_3$, $E_1 + E_2 + E_3 = I$ y $E_iE_j = 0$, $i \neq j$.

7. En los Ejercicios 5 y 6, obsérvese que (para todo i) el espacio de vectores propios asociados con el valor propio c_i es generado por los vectores columna de las matrices E_j con $j \neq i$. ¿Es una coincidencia?

8. Sea T un operador lineal sobre V que conmuta con todo operador proyección sobre V. ¿Que se puede decir de T?

9. Sea V el espacio vectorial de las funciones reales continuas sobre el intervalo $[-1, 1]$ del eje real. Sea W_p el subespacio de las funciones pares, $f(-x) = f(x)$, y sea W_i el subespacio de las funciones impares, $f(-x) = -f(x)$.

(a) Demostrar que $V = W_p \oplus W_i$.

(b) Si T es el operador integración indefinida

$$(Tf)(x) = \int_0^x f(t)\, dt$$

¿son W_p y W_i invariantes por T?

6.8. Teorema de descomposición prima

Estamos tratando de estudiar un operador lineal T sobre el espacio V de dimensión finita, por descomposición de T en suma directa de operadores que son en cierto sentido elementales. Esto se puede hacer, en ciertos casos especiales, por medio de los valores propios y los vectores propios de T; es decir, cuando el polinomio minimal de T se puede factorizar sobre el cuerpo de los escalares F como producto de polinomios mónicos distintos de grado 1. ¿Qué se puede hacer en el caso del T general? Si se estudia T usando los valores propios, tenemos dos problemas. Primero, T puede no tener un valor propio simple; esto es realmente una deficiencia del campo escalar, a saber: que no es algebraicamente cerrado. Segundo, incluso si el polinomio característico se puede factorizar completamente sobre F, como producto de polinomios de grado 1, puede ser que no haya suficientes vectores propios para T que generen el espacio V; esto es, obviamente, una deficiencia de T. La segunda situación queda ilustrada por el operador T en F^3 (F cualquier cuerpo) representado en la base canónica por

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

El polinomio característico para A es $(x - 2)^2(x + 1)$ y éste es también el polinomio minimal para A (o para T). Así que T no es diagonalizable. Se ve que esto sucede porque el espacio nulo de $(T - 2I)$ tiene solo dimensión 1. Por otro lado, el espacio nulo de $(T + I)$ y el espacio nulo de $(T - 2I)^2$, juntos, generan V, siendo el primero, el subespacio generado por ϵ_3 y el último, el subespacio generado por ϵ_1 y ϵ_2.

Este será más o menos el método general para el segundo problema. Si (recuérdese que es una suposición) el polinomio minimal de T se descompone en

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}$$

donde $c_1, \ldots, c_k$ son elementos distintos de F, entonces se verá que el espacio V es suma directa de los espacios nulos de $(T - c_iI)^{r_i}$, $i = 1, \ldots, k$. La hipó-

tesis para p es equivalente al hecho que T es triangulable (Teorema 5); sin embargo, ese conocimiento no va a ayudar.

El teorema que se probará es más general de lo que se ha dicho, ya que es aplicable para la descomposición prima del polinomio minimal, sean o no sean de primer grado todos los factores primos. Le será de provecho al lector pensar en el caso especial cuando los factores primos son de grado 1, y aún más particularmente pensar en la demostración del tipo proyección del Teorema 6, un caso especial de este teorema.

Teorema 12 (teorema de la descomposición prima). *Sea T un operador lineal sobre el espacio vectorial V de dimensión finita sobre el cuerpo F. Sea p el polinomio minimal de T,*

$$p = p_1^{r_1} \cdots p_k^{r_k}$$

donde los p_i son polinomios mónicos irreducibles distintos sobre F, y los r_i son enteros positivos. Sea W_i el espacio nulo de $p_i(T)^{r_i}$, $i = 1, \ldots, k$. Entonces

(i) $V = W_1 \oplus \cdots \oplus W_k$;

(ii) *cada E_i es invariante por T;*

(iii) *si T_i es el operador inducido sobre W_i por T, entonces el polinomio minimal de T_i es $p_i^{r_i}$.*

Demostración. La idea de la demostración es la siguiente. Si la descomposición en suma directa (i) es válida, ¿cómo se pueden determinar las proyecciones $E_1, \ldots, E_k$ asociadas a la descomposición? La proyección E_i será la identidad sobre W_i y cero sobre los otros W_j. Se encontrará un polinomio h_i tal que $h_i(T)$ es la identidad sobre W_i y es cero sobre los otros W_j, con lo que $h_1(T) + \cdots + h_k(T) = I$, etc.

Para todo i, sea

$$f_i = \frac{p}{p_i^{r_i}} = \prod_{j \neq i} p_j^{r_j}.$$

Como $p_1, \ldots, p_k$ son polinomiales primos distintos, los polinomios $f_1, \ldots, f_k$ son primos relativos (Teorema 10, Capítulo 4). Así que existen polinomios $g_1, \ldots, g_k$ tales que

$$\sum_{i=1}^{n} f_i g_i = 1.$$

Nótese también que si $i \neq j$, entonces $f_i f_j$ es divisible por el polinomio p, pues $f_i f_j$ contiene a cada $p_m^{r_m}$ como factor. Se demostrará que los polinomios $h_i = f_i g_i$ se comportan del mismo modo descrito al comienzo de la demostración.

Sea $E_i = h_i(T) = f_i(T)g_i(T)$. Como $h_1 + \cdots + h_k = 1$ y p divide a $f_i f_j$ para $i \neq j$, se tiene

$$E_1 + \cdots + E_k = I$$
$$E_i E_j = 0, \qquad \text{si } i \neq j.$$

Con lo que E_i son proyecciones que corresponden a una descomposición en suma directa del espacio V. Deseamos ahora hacer ver que la imagen de E_i es

exactamente el subespacio W_i. Es claro que cada vector de la imagen de E_i está en W_i; en efecto, si α está en la imagen de E_i, entonces $\alpha = E_i\alpha$ y así

$$\begin{aligned} p_i(T)^{r_i}\alpha &= p_i(T)^{r_i}E_i\alpha \\ &= p_i(T)^{r_i}f_i(T)g_i(T)\alpha \\ &= 0 \end{aligned}$$

pues $p^{r_i}f_ig_i$ es divisible por el polinomio minimal p. Recíprocamente, supóngase que α está en el espacio nulo de $p_i(T)^{r_i}$. Si $j = i$, entonces f_jg_j es divisible por $p_i^{r_i}$, con lo que $f_j(T)g_j(T)\alpha = 0$, es decir, $E_j\alpha = 0$ para $j \neq i$. Pero entonces es inmediato que $E_i\alpha = \alpha$, es decir, que α está en la imagen de E_i. Esto completa la demostración de la parte (i) de la tesis.

Es claro que los subespacios W_i son invariantes por T. Si T_i es el operador inducido en W_i por T, entonces evidentemente $p_i(T_i)^{r_i} = 0$, ya que, por definición, $p_i(T)^{r_i} = 0$ en el subespacio W_i. Esto muestra que el polinomio minimal de T_i divide a $p_i^{r_i}$. Recíprocamente, sea g cualquier polinomio tal que $g(T_i) = 0$. Entonces $g(T)f_i(T) = 0$. Con lo que gf_i es divisible por el polinomio minimal p de T; es decir, $p_i^{r_i}f_i$ divide a gf_i. Es fácil ver que $p_i^{r_i}$ divide a g. Luego el polinomio minimal de T_i es $p_i^{r_i}$. ∎

Corolario. *Si $E_1, \ldots, E_k$ son las proyecciones asociadas a la descomposición prima de T, entonces todo E_i es un polinomio de T, y en consecuencia, si un operador lineal U conmuta con T, entonces U conmuta con cada uno de los E_i; es decir, cada subespacio W_i es invariante por U.*

Con la notación de la demostración del Teorema 12, consideremos el caso especial en que el polinomio minimal de T es un producto de polinomios de primer grado; es decir, el caso en que cada p_i es de la forma $p_i = x - c_i$. Ahora la imagen de E_i es el espacio nulo W_i de $(T - c_iI)^{r_i}$. Hágase $D = c_1E_1 + \cdots + c_kE_k$. Por el Teorema 11, D es un operador diagonalizable que se llama la **parte diagonalizable** de T. Considérese el operador $N = T - D$. Ahora

$$\begin{aligned} T &= TE_1 + \cdots + TE_k \\ D &= c_1E_1 + \cdots + c_kE_k \end{aligned}$$

así

$$N = (T - c_1I)E_1 + \cdots + (T - c_kI)E_k.$$

El lector debe estar ya lo suficientemente familiarizado con las proyecciones para ver ahora que

$$N^2 = (T - c_1I)^2E_1 + \cdots + (T - c_kI)^2E_k$$

y, en general, que

$$N^r = (T - c_1I)^rE_1 + \cdots + (T - c_kI)^rE_k.$$

Cuando $r \geq r_i$ para cada i, se tendrá que $N^r = 0$, ya que el operador $(T - c_iI)^r$ será entonces 0 en el recorrido de E_i.

Definición. *Sea N un operador lineal sobre el espacio vectorial V. Se dice que N es* **nilpotente** *si existe un entero positivo r tal que $N^r = 0$.*

Teorema 13. *Sea T un operador lineal sobre el espacio vectorial V de dimensión finita sobre el cuerpo F. Supóngase que el polinomio minimal de T se descompone sobre F, en producto de polinomios lineales. Entonces existen un operador diagonalizable D sobre V y un operador N nilpotente sobre V tales que*

(i) $T = D + N$,
(ii) $DN = ND$.

El operador diagonalizable D y el operador nilpotente N están unívocamente determinados por (i) *y* (ii), *y cada uno de ellos es un polinomio de T.*

Demostración. Acabamos de observar que se puede escribir $T = D + N$, donde D es diagonalizable y N nilpotente, y donde D y N no solo conmutan, sino que son polinomios de T. Supóngase ahora que también tenemos $T = D' + N'$, donde D' es diagonalizable, N' nilpotente y $D'N' = N'D'$. Demostraremos que $D = D'$ y $N = N'$.

Como D' y N' conmutan entre sí y $T = D' + N'$, se ve que D' y N' conmutan con T. Así, D' y N' conmutan con cualquier polinomio de T; luego conmutan con D y con N. Ahora tenemos

$$D + N = D' + N'$$

o

$$D - D' = N' - N$$

y todos estos cuatro operadores conmutan entre sí. Como D y D' son ambos diagonalizables y conmutan, son simultáneamente diagonalizables, y $D - D'$ es diagonalizable. Como N y N' son nilpotentes y conmutan, el operador $(N' - N)$ es nilpotente; en efecto, como N y N' conmutan

$$(N' - N)^r = \sum_{j=0}^{r} \binom{r}{j} (N')^{r-j}(-N)^j$$

y así, cuando r es suficientemente grande, cada término en esta expresión de $(N' - N)^r$ será 0. (En realidad un operador nilpotente en un espacio de dimensión n debe tener su potencia n-ésima 0; si se hace arriba $r = 2n$, será suficientemente grande. Se sigue entonces que $r = n$ es suficientemente grande, pero esto no es obvio en la expresión anterior.) Ahora $D - D'$ es un operador diagonalizable y también nilpotente. Un tal operador es evidentemente el operador cero, pues como es nilpotente, el polinomio minimal de este operador es de la forma x^r para algún $r \leq m$; pero entonces, como el operador es diagonalizable, el polinomio mínimal no puede tener raíces repetidas; luego $r = 1$ y el polinomio minimal es simplemente x, lo que dice que el operador es 0. Con lo que se tiene que $D = D'$ y $N = N'$. ∎

Corolario. *Sea V un espacio vectorial de dimensión finita sobre un cuerpo F algebraicamente cerrado, v. g., el cuerpo de los números complejos. Entonces todo operador lineal T sobre V puede escribirse como suma de un operador diagonalizable D y un operador nilpotente N que conmutan. Estos operadores D y N son únicos y cada uno es un polinomio de T.*

Por estos resultados se ve que el estudio de los operadores lineales en un espacio vectorial sobre un cuerpo algebraicamente cerrado queda esencialmente reducido al estudio de operadores nilpotentes. Para espacios vectoriales sobre cuerpos no algebraicamente cerrados se debe encontrar todavía un sustituto de los valores y vectores propios. Es un hecho muy interesante que estos dos problemas pueden tratarse en forma simultánea, y ello es lo que se hará en el próximo capítulo.

Para concluir esta sección queremos dar un ejemplo que ilustre algunas de las ideas del teorema de descomposición prima. Hemos preferido darlo al final de la sección, ya que se refiere a ecuaciones diferenciales y, por tanto, no es exclusivamente de álgebra lineal.

Ejemplo 14. En el teorema de descomposición prima no es necesario que el espacio vectorial V sea de dimensión finita, ni es necesario para las partes (i) y (ii) que p sea el polinomio minimal de T. Si T es un operador lineal sobre un espacio vectorial arbitrario y si existe un polinomio mónico p, tal que $p(T) = 0$, entonces las partes (i) y (ii) del Teorema 12 son válidas para T con la demostración que se ha dado.

Sea n un entero positivo y sea V el espacio de todas las funciones f n veces continuamente derivables sobre el eje real y que satisfacen la ecuación diferencial

$$\text{(6-18)} \qquad \frac{d^n f}{dt^n} + a_{n-1}\frac{d^{n-1}f}{dt^{n-1}} + \cdots + a_1\frac{df}{dt} + a_0 f = 0$$

donde los $a_0, \ldots, a_{n-1}$ son constantes dadas. Si C_n representa el espacio de las funciones n veces continuamente derivables, entonces el espacio V de las soluciones de esta ecuación diferencial es un subespacio de C_n. Si D representa el operador derivación y p es el polinomio

$$p = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

entonces V es el espacio nulo del operador $p(D)$, ya que (6-18) no dice más que $p(D)f = 0$. Por tanto, V es invariante por D. Considérese ahora D como un operador lineal sobre el subespacio V. Entonces $p(D) = 0$.

Si estamos considerando funciones derivables de valores complejos, entonces C_n y V son espacios vectoriales complejos, y $a_0, \ldots, a_{n-1}$ puede ser cualquier número complejo. Escribimos ahora

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}$$

donde los $c_1, \ldots, c_k$ son números complejos distintos. Si W_j es el espacio nulo de $(D - c_jI)^{r_j}$, entonces el Teorema 12 dice que

$$V = W_1 \oplus \cdots \oplus W_k.$$

Es decir, si f satisface la ecuación diferencial (6-18), entonces f queda unívocamente determinado en la forma

$$f = f_1 + \cdots + f_k$$

donde las f_j satisfacen la ecuación diferencial $(D - c_jI)^{r_j}f_j = 0$. Así el estudio

de las soluciones de la ecuación (6-18) queda reducido al del espacio de soluciones de una ecuación diferencial de la forma

(6-19) $$(D - cI)^r f = 0.$$

Esta reducción se ha efectuado por los métodos generales del álgebra lineal, es decir, por el teorema de descomposición prima.

Para describir el espacio de soluciones de (6-19), se debe conocer algo respecto a las ecuaciones diferenciales; esto es, se debe conocer algo respecto a D, además del hecho de que es un operador lineal. Es muy fácil demostrar, por inducción sobre r, que si f está en C_r, entonces

$$(D - cI)^r f = e^{ct} D^r(e^{-ct} f)$$

esto es,

$$\frac{df}{dt} - cf(t) = e^{ct} \frac{d}{dt}(e^{-ct} f), \quad \text{etc.}$$

Así, $(D - cI)^r f = 0$ si, y solo si, $D^r(e^{-ct} f) = 0$. Una función g tal que $D^r g = 0$, es decir, $d^r g/dt^r = 0$, debe ser una función polinomio de grado $(r - 1)$ o menor

$$g(t) = b_0 + b_1 t + \cdots + b_{r-1} t^{r-1}.$$

Así que f satisface (6-19) si, y solo si, tiene la forma

$$f(t) = e^{ct}(b_0 + b_1 t + \cdots + b_{r-1} t^{r-1}).$$

En consecuencia, las «funciones» e^{ct}, te^{ct}, ..., $t^{r-1}e^{ct}$ generan el espacio de soluciones de (6-19). Como 1, t, ..., t^{r-1} son funciones linealmente independientes, y la función exponencial no tiene ceros, estas r funciones $t^j e^{ct}$, $0 \leq j \leq r - 1$, forman una base del espacio de soluciones.

Volviendo a la ecuación diferencial (6-18), que es

$$p(D)f = 0$$
$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}$$

se ve que las n funciones $t^m e^{c_j t}$, $0 \leq m \leq r_j - 1$, $1 \leq j \leq k$, forman una base del espacio de soluciones de (6-18). En particular, el espacio de soluciones es de dimensión finita y tiene dimensión igual al grado del polinomio p.

Ejercicios

1. Sea T un operador lineal sobre R^2 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{bmatrix}.$$

Expresar el polinomio minimal p de T en la forma $p = p_1 p_2$, donde p_1 y p_2 son polinomios mónicos e irreducibles sobre el cuerpo de los números reales. Sea W_i el espacio nulo de $p_i(T)$. Hallar las bases $\mathcal{B}_i$ para los espacios W_1 y W_2. Si T_i es el operador inducido en W_i por T, hallar la matriz de T_i en la base $\mathcal{B}_i$ (anteriormente citada).

2. Sea T el operador lineal sobre R^3 representado por la matriz

$$\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$$

en la base ordenada canónica. Demostrar que existen un operador diagonalizable D sobre R^3 y un operador nilpotente N sobre R^3 tales que $T = D + N$ y $DN = ND$. Hallar las matrices de D y N en la base canónica. (No hay más que repetir la demostración del Teorema 12 para este caso especial.)

3. Si V es el espacio de los polinomios de grado menor o igual que n sobre un cuerpo F, demostrar que el operador derivación sobre V es nilpotente.

4. Sea T el operador lineal sobre el espacio V de dimensión finita con polinomio característico

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

y polinomio minimal

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}.$$

Sea W_i el espacio nulo de $(T - c_iI)^{r_i}$.

(a) Demostrar que W_i es el conjunto de todos los vectores α de V tales que $(T - c_iI)^m\alpha = 0$ para algún entero positivo m (que dependerá de α).

(b) Demostrar que la dimensión de W_i es d_i. (*Sugerencia*: Si T_i es el operador inducido en W_i por T, entonces $T_i - c_iI$ es nilpotente; así el polinomio característico de $T_i - c_iI$ debe ser x^{e_i}, donde e_i es la dimensión de W_i (¿demostración?); así el polinomio característico de T_i es $(x - c_i)^{e_i}$; ahora úsese el hecho de que el polinomio característico de T es el producto de los polinomios característicos de los T_i, para demostrar que $e_i = d_i$.)

5. Sea V un espacio vectorial de dimensión finita sobre el cuerpo de los números complejos. Sea T un operador lineal sobre V y sea D la parte diagonalizable de T. Demostrar que si g es cualquier polinomio con coeficientes complejos, entonces la parte diagonalizable de $g(T)$ es $g(D)$.

6. Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea T un operador lineal sobre T tal que rango $(T) = 1$. Demostrar que o bien T es diagonalizable o bien T es nilpotente, pero no ambas cosas simultáneamente.

7. Sea V un espacio vectorial de dimensión finita y sea T un operador lineal sobre V. Supóngase que T conmuta con cada operador lineal diagonalizable en V. Demostrar que T es un múltiplo escalar del operador identidad.

8. Sea V el espacio de las matrices $n \times n$ sobre el cuerpo F y sea A una matriz dada $n \times n$ sobre F. Se define un operador lineal T sobre V por $T(B) = AB = BA$. Demostrar que si A es una matriz nilpotente, T es un operador nilpotente.

9. Dar un ejemplo de dos matrices nilpotentes 4×4 con el mismo polinomio minimal (necesariamente tienen el mismo polinomio característico), pero que no son semejantes.

10. Sea T un operador lineal sobre el espacio V de dimensión finita, sea $p = p_1^{e_1} \cdots p_k^{e_k}$ el polinomio minimal de T y sea $V = W_1 \oplus \cdots \oplus W_k$ la descomposición prima de T; es decir, W_j es el espacio nulo de $p_j(T)^{r_j}$. Sea W cualquier subespacio de V invariante por T. Demostrar que

$$W = (W \cap W_1) \oplus (W \cap W_2) \oplus \cdots \oplus (W \cap W_k).$$

11. ¿Cuál es el error en la siguiente demostración del Teorema 13? Supóngase que el polinomio minimal de T sea un producto de factores lineales. Entonces, por el Teorema 5, T es triangulable. Sea $\mathcal{B}$ una base ordenada tal que $A = [T]_{\mathcal{B}}$ es triangular superior. Sea D la matriz diagonal con elementos en la diagonal principal $a_{11}, \ldots, a_{nn}$. Entonces $A = D + N$, donde N es estrictamente triangular superior. Evidentemente, N es nilpotente.

12. Si se meditó en el Ejercicio 11, pensar nuevamente en él, después de observar qué dice el Teorema 7 respecto a las partes diagonalizable y nilpotente de T.

13. Sea T un operador lineal V con polinomio minimal de la forma p^n con p irreducible sobre el cuerpo de los escalares. Demostrar que existe un vector α en V tal que el T-anulador de α es p^n.

14. Usar el teorema de descomposición prima y el resultado del Ejercicio 13 para demostrar lo siguiente. Si T es cualquier operador lineal sobre un espacio V de dimensión finita, entonces existe un vector α en V con T-anulador igual al polinomio minimal de T.

15. Si N es un operador lineal nilpotente sobre un espacio de dimensión n, entonces el polinomio característico de N es x^n.

7. Las formas racional y de Jordan

7.1. Subespacios cíclicos y anuladores

Sea nuevamente V un espacio vectorial de dimensión finita sobre un cuerpo F y sea T un operador lineal dado (pero arbitrario) sobre V. Si α es cualquier vector de V, existe un subespacio mínimo de V que es invariante por T y que contiene a α. Este subespacio se puede definir como la intersección de todos los subespacios invariantes por T que contienen a α; sin embargo, es más útil por el momento ver las cosas del siguiente modo. Si W es un subespacio de V invariante por T y que contiene a α, entonces W debe contener también al vector $T\alpha$; luego W debe contener a $T(T\alpha) = T^2\alpha$, $T(T^2\alpha) = T^3\alpha$, etc. Esto es, W debe contener a $g(T)\alpha$ para todo polinomio g sobre F. El conjunto de todos los vectores de la forma $g(T)\alpha$, con g en $F[x]$, es evidentemente invariante por T, y es así el menor subespacio invariante por T que contiene a α.

Definición. *Si α es cualquier vector de V, el* **subespacio T-cíclico generado por** *α es el subespacio $Z(\alpha; T)$ de los vectores de la forma $g(T)\alpha$, g en $F[x]$. Si $Z(\alpha; T) = V$ entonces se dice que α es un* **vector cíclico** *de T.*

Otro modo de describir el subespacio $Z(\alpha; T)$ es que $Z(\alpha; T)$ es el subespacio generado por los vectores $T^k\alpha$, $k \geq 0$; y así α es un vector cíclico de T si, y solo si, estos vectores generan V. Se previene al lector que el operador general T no tiene vectores cíclicos.

Ejemplo 1. Para cualquier T, el subespacio T-cíclico generado por el vector nulo es el subespacio nulo. El espacio $Z(\alpha; T)$ es de dimensión uno si, y solo si, α es un vector propio de T. Para el operador identidad, todo vector no nulo genera un subespacio cíclico unidimensional; así, si dim $V > 1$, el operador

identidad no tiene vector cíclico. Un ejemplo de operador que tiene un vector cíclico es el operador lineal T sobre F^2 que está representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Aquí el vector cíclico (un vector cíclico) es ϵ_1; en efecto, si $\beta = (a, b)$, entonces con $g = a + bx$ se tiene $\beta = g(T)\epsilon_1$. Para este mismo operador el subespacio cíclico generado por ϵ_2 es el espacio unidimensional generado por ϵ_2, pues ϵ_2 es un vector propio de T.

Para cualquier T y α sean las relaciones lineales

$$c_0\alpha + c_1T\alpha + \cdots + c_kT^k\alpha = 0$$

entre los vectores $T^j\alpha$, esto es, se consideran los polinomios $g = c_0 + c_1x + \cdots + c_kx^k$ que tienen la propiedad de que $g(T)\alpha = 0$. El conjunto de todos los g en $F[x]$ tales que $g(T)\alpha = 0$ es claramente un ideal en $F[x]$. Es también un ideal no nulo, ya que contiene al polinomio minimal p del operador $T(p(T)\alpha = 0$ para todo α de V).

Definición. *Si α es cualquier vector de V, el* **T-anulador** *de α es el ideal $M(\alpha; T)$ en $F[x]$ que consta de todos los polinomios g sobre F de modo que $g(T)\alpha = 0$. El polinomio mónico único p_α que genera este ideal se le llamará también el* **T-anulador** *de α.*

Como se observó anteriormente, el T-anulador p_α divide al polinomio minimal del operador T. El lector debe observar también que grd $(p_\alpha) > 0$, salvo que α sea el vector cero.

Teorema 1. *Sea α un vector no nulo en V y sea p_α el T-anulador de α.*

(i) *El grado de p_α es igual a la dimensión del subespacio cíclico $Z(\alpha; T)$.*

(ii) *Si el grado de p_α es k, entonces los vectores $\alpha, T\alpha, T^2\alpha, \ldots, T^{k-1}\alpha$ forman una base de $Z(\alpha; T)$.*

(iii) *Si U es el operador lineal en $Z(\alpha; T)$ inducido por T, entonces el polinomio minimal de U es p_α.*

Demostración. Sea g un polinomio sobre el cuerpo F. Se escribe

$$g = p_\alpha q + r$$

donde $r = 0$ o grd $(r) <$ grd $(p_\alpha) = k$. El polinomio $p_\alpha q$ está en el T-anulador de α, y así

$$g(T)\alpha = r(T)\alpha.$$

Como $r = 0$ o grd $(r) < k$, el vector $r(T)\alpha$ es combinación lineal de los vectores $\alpha, T\alpha, \ldots, T^{k-1}\alpha$, y como $g(T)\alpha$ es un vector típico en $Z(\alpha; T)$, esto muestra que estos k vectores generan $Z(\alpha; T)$. Estos vectores son, en efecto, linealmente independientes, ya que cualquier relación lineal no trivial entre ellos dará un

polinomio g no nulo tal que $g(T)\alpha = 0$ y grd $(g) <$ grd (p_α), que es absurdo. Esto demuestra (i) y (ii).

Sea U el operador lineal en $Z(\alpha; T)$ que se obtiene por restricción de T a ese subespacio. Si g es un polinomio sobre F, entonces

$$\begin{aligned} p_\alpha(U)g(T)\alpha &= p_\alpha(T)g(T)\alpha \\ &= g(T)p_\alpha(T)\alpha \\ &= g(T)0 \\ &= 0. \end{aligned}$$

Así, el operador $p_\alpha(U)$ aplica cada vector de $Z(\alpha; T)$ en 0 y es el operador cero en $Z(\alpha; T)$. Además, si h es un polinomio de grado menor que k, no se puede tener $h(U) = 0$; en efecto, entonces $h(U)\alpha = h(T)\alpha = 0$, que contradice la definición de p_α. Esto demuestra que p_α es el polinomio minimal de U. ∎

Una consecuencia particular de este teorema es la siguiente. Si sucede que α es un vector cíclico para T, entonces el polinomio minimal de T debe tener igual grado que la dimensión del espacio V; luego el teorema de Cayley-Hamilton dice que el polinomio minimal de T es el polinomio característico de T. Se demostrará más adelante que para cualquier T existe un vector α en V que tiene al polinomio minimal de T por anulador. Se deducirá entonces que T tiene un vector cíclico si, y solo si, los polinomios minimal y característico de T son idénticos. Pero ello tomará algún trabajo antes de poderlo ver.

El plan es estudiar el T general usando operadores que tienen un vector cíclico. Para ello se considera un operador lineal U, sobre el espacio V de dimensión k, que tiene un vector cíclico α. Por el Teorema 1 los vectores $\alpha, \ldots, U^{k-1}\alpha$ forman una base del espacio W, y el anulador p_α de α es el polinomio minimal de U (y entonces, también el polinomio característico de U). Si se hace $\alpha_i = U^{i-1}\alpha$, $i = 1, \ldots, k$, entonces la acción de U sobre la base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_k\}$ es

$$\text{(7-1)} \qquad \begin{aligned} U\alpha_i &= \alpha_{i+1}, \qquad i = 1, \ldots, k-1 \\ U\alpha_k &= -c_0\alpha_1 - c_1\alpha_2 - \cdots - c_{k-1}\alpha_k \end{aligned}$$

donde $p = c_0 + c_1x + \cdots + c_{k-1} + x^k$. La expresión para $U\alpha_k$ se desprende del hecho de que $p_\alpha(U)\alpha = 0$, es decir,

$$U^k\alpha + c_{k-1}U^{k-1}\alpha + \cdots + c_1U\alpha + c_0\alpha = 0.$$

Esto dice que la matriz de U, en la base ordenada $\mathcal{B}$, es

$$\text{(7-2)} \qquad \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -c_{k-1} \end{bmatrix}.$$

La matriz (7-2) se llama la **matriz asociada** al polinomio mónico p_α.

Teorema 2. *Si U es un operador lineal sobre un espacio de dimensión finita W, entonces U tiene un vector cíclico si, y solo si, existe una base ordenada*

de W en la que U esté representada por la matriz asociada al polinomio minimal de U.

Demostración. Antes se observó que si U tiene un vector cíclico, entonces existe tal base ordenada de W. Recíprocamente, si se tiene una base ordenada $\{\alpha_1, \ldots, \alpha_k\}$ de W en la que U esté representada por la matriz asociada a su polinomio minimal, es obvio que α_1 es un vector cíclico de U. ∎

Corolario. *Si A es la matriz asociada a un polinomio mónico p, entonces p es el polinomio minimal y el polinomio característico de A.*

Demostración. Una manera de demostrarlo es hacer que U sea el operador lineal sobre F^k, representado por A' en la base ordenada canónica, y aplicar el Teorema 1 junto con el teorema de Cayley-Hamilton. Otro método consiste en usar el Teorema 1 para ver que p es el polinomio minimal de A y verificar por un cálculo directo que p es el polinomio característico de A. ∎

Un último comentario: si T es un operador lineal cualquiera sobre el espacio V y α es un vector de V, entonces el operador U que T induce en el subespacio cíclico $Z(\alpha; T)$ tiene un vector cíclico, que es α. Así, $Z(\alpha; T)$ tiene una base ordenada en la cual U está representado por la matriz asociada de p_α, T-anulador de α.

Ejercicios

1. Sea T un operador lineal sobre F^2. Demostrar que un vector no nulo, que no es un vector propio de T, es un vector cíclico de T. Demostrar luego que o T tiene un vector cíclico o T es un múltiplo escalar del operador identidad.

2. Sea T el operador lineal sobre R^3 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Demostrar que T no tiene vector cíclico. ¿Cuál es el subespacio T-cíclico generado por el vector $(1, -1, 3)$?

3. Sea T el operador lineal sobre C^3 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 1 & i & 0 \\ -1 & 2 & -i \\ 0 & 1 & 1 \end{bmatrix}.$$

Hallar el T-anulador del vector $(1, 0, 0)$. Hallar el T-anulador de $(1, 0, i)$.

4. Demostrar que si T^2 tiene un vector cíclico, entonces T tiene un vector cíclico. ¿Es verdadero el recíproco?

5. Sea V un espacio vectorial de dimensión n sobre el cuerpo F y sea N un operador lineal

nilpotente sobre V. Supóngase que $N^{n-1} \neq 0$ y sea α un vector cualquiera de V de modo que $N^{n-1}\alpha \neq 0$. Demostrar que α es un vector cíclico de N. ¿Cuál es exactamente la matriz de N en la base ordenada $\{\alpha, N\alpha, \ldots, N^{n-1}\alpha\}$?

6. Dar una demostración directa de que si A es la matriz asociada al polinomio mónico p, entonces p es el polinomio característico de A.

7. Sea V un espacio vectorial de dimensión n y sea T un operador lineal sobre V. Supóngase que T es *diagonalizable*.

(a) Si T tiene un vector cíclico, demostrar que T tiene n valores propios distintos.

(b) Si T tiene n valores propios distintos, y si $\{\alpha_1, \ldots, \alpha_n\}$ es una base de vectores propios de T, demostrar que $\alpha = \alpha_1 + \cdots + \alpha_n$ es un vector cíclico de T.

8. Sea T un operador lineal sobre el espacio vectorial V de dimensión finita. Supóngase que T tiene un vector cíclico. Demostrar que si U es un operador lineal cualquiera que conmute con T, U es un polinomio de T.

7.2. Descomposiciones cíclicas y forma racional

El objeto de esta sección es demostrar que si T es un operador lineal arbitrario sobre un espacio V de dimensión finita, entonces existen vectores $\alpha_1, \ldots, \alpha_r$ en V tales que

$$V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T).$$

En otras palabras, queremos demostrar que V es una suma directa de subespacios T-cíclicos, lo cual mostrará que T es la suma directa de un número finito de operadores lineales, cada uno de los cuales tiene un vector cíclico. El efecto de esto será reducir muchos problemas acerca del operador lineal general a problemas análogos con un operador que tiene un vector cíclico. El teorema que se demostrará (Teorema 3) es uno de los resultados más profundos del álgebra lineal y tiene muchos corolarios interesantes.

El teorema de la descomposición cíclica está estrechamente relacionado con el siguiente problema, ¿Qué subespacios W, invariantes por T, tienen la propiedad de que existe un subespacio W', invariante por T, tal que $V = W \oplus W'$? Si W es cualquier subespacio de un espacio V de dimensión finita, entonces existe un subespacio W' tal que $V = W \oplus W'$. En general hay muchos de tales subespacios W' y cada uno de ellos se llama **complementario** de W. Se pregunta, ¿cuándo un subespacio invariante por T tiene un subespacio complementario también invariante por T?

Supóngase que $V = W \oplus W'$, donde W y W' son invariantes por T, y véase entonces qué se puede descubrir respecto al subespacio W. Cada vector β de V es de la forma $\beta = \gamma + \gamma'$, donde γ está en W y γ' en W'. Si f es un polinomio sobre el cuerpo de los escalares, entonces

$$f(T)\beta = f(T)\gamma + f(T)\gamma'.$$

Como W y W' son invariantes por T, el vector $f(T)\gamma$ está en W y $f(T)\gamma'$ está en W'. Por tanto, $f(T)\beta$ está en W si, y solo si, $f(T)\gamma' = 0$. Lo que interesa es el hecho, aparentemente sin importancia, de que si $f(T)\beta$ está en W, entonces $f(T)\beta = f(T)\gamma$.

Definición. *Sea T un operador lineal sobre un espacio vectorial V y sea W un subespacio de V. Se dice que W es T-***admisible** *si*

(i) *W es invariante por T;*

(ii) *si $f(T)\beta$ está en W, existe un vector γ en W tal que $f(T)\beta = f(T)\gamma$.*

Como acabamos de mostrar, si W es invariante y tiene un subespacio invariante complementario, entonces W es admisible. Una de las consecuencias del Teorema 3 será el recíproco, de modo que la admisibilidad caracterizará aquellos subespacios invariantes que tienen subespacios invariantes complementarios.

Indiquemos cómo la propiedad de la admisibilidad está implicada en la intención de obtener una descomposición

$$V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T).$$

El método básico para llegar a tal descomposición será seleccionar los vectores $\alpha_1, \ldots, \alpha_r$ por inducción. Supóngase que, por un proceso u otro, se hayan seleccionado $\alpha_1, \ldots, \alpha_j$, y que

$$W_j = Z(\alpha_1; T) + \cdots + Z(\alpha_j; T)$$

sea un subespacio propio. Se desea encontrar un vector no nulo $\alpha_j + 1$ tal que

$$W_j \cap Z(\alpha_{j+1}; T) = \{0\}$$

porque el subespacio $W_{j+1} = W_j \oplus Z(\alpha_{j+1}; T)$ será entonces al menos una dimensión más cercana al espacio V que se descompone. Pero ¿por qué debe existir tal vector α_{j+1}? Si $\alpha_1, \ldots, \alpha_j$, han sido elegidos de modo que W_j sea un subespacio T-admisible, entonces es relativamente fácil ver que se puede encontrar un α_{j+1} adecuado. Esto es lo que hará posible la demostración del Teorema 3, aun cuando no expresemos así el razcnamiento fraseado.

Sea W un subespacio propio invariante por T. Se desea encontrar un vector α, no nulo, tal que

$$W \cap Z(\alpha; T) = \{0\}. \tag{7-3}$$

Se puede elegir un vector β que no está en W. Considérese el *T-conductor* $S(\beta; W)$, que consta de todos los polinomios g tales que $g(T)\beta$ está en W. Se recuerda que el polinomio mónico $f = s(\beta; W)$ que genera el ideal $S(\beta; W)$ es llamado también el *T-conductor de β en W*. El vector $f(T)\beta$ está en W. Ahora, si W es T-admisible, existe un γ en W con $f(T)\beta = f(T)\gamma$. Sea $\alpha = \beta - \gamma$ y sea g un polinomio cualquiera. Como $\beta - \gamma$ está en W, $g(T)\beta$ estará en W si, y solo si, $g(T)\alpha$ está en W; en otras palabras, $S(\alpha; W) = S(\beta; W)$. Con lo que el polinomio f es también el T-conductor de α en W. Pero $f(T)\alpha = 0$. Ello dice que $g(T)\alpha$ está en W si, y solo si, $g(T)\alpha = 0$; es decir, los subespacios $Z(\alpha; T)$ y W son linealmente independientes (7-3) y f es el T-anulador de α.

Teorema 3 (teorema de descomposición cíclica). *Sea T un operador lineal sobre un espacio vectorial de dimensión finita V y sea W_0 un subespacio propio T-admisible de V. Existen vectores no nulos $\alpha_1, \ldots, \alpha_r$ en V con T-anuladores respectivos $p_1, \ldots, p_r$, tales que*

(i) $V = W_0 \oplus Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T)$;
(ii) p_k *divide a* p_{k-1}, $k = 2, \ldots, r$.

Más aún, el entero r y los anuladores $p_1, \ldots, p_r$ están unívocamente determinados por (i), (ii) *y el hecho de que ninguno de los α_k es* 0.

Demostración. La demostración es bastante larga, razón por la cual la dividiremos en cuatro partes. Para la primera lectura parece más fácil tomar $W_0 = \{0\}$, aun cuando no produzca una simplificación sustancial. En toda la demostración se abreviará $f(T)\beta$ por $f\beta$.

Parte 1. *Existen vectores no nulos $\beta_1, \ldots, \beta_r$ en V tales que*

(a) $V = W_0 + Z(\beta_1; T) + \cdots + Z(\beta_r; T)$;
(b) *si* $1 \leq k \leq r$ *y*

$$W_k = W_0 + Z(\beta_1; T) + \cdots + Z(\beta_k; T)$$

entonces el conductor $p_k = s(\beta_k; W_{k-1})$ tiene el grado máximo entre todos los T-conductores en el subespacio W_{k-1}; es decir, para todo k

$$grd\, p_k = \max_{\alpha \text{ en } V} grd\, s(\alpha; W_{k-1})$$

Esta parte depende solo del hecho de que W_0 es un subespacio invariante. Si W es un subespacio propio invariante por T, entonces

$$0 < \max_{\alpha} \operatorname{grd} s(\alpha; W) \leq \dim V$$

y se puede elegir un vector β de modo que grd $s(\beta; W)$ alcance tal máximo. El subespacio $W + Z(\beta; T)$ es entonces invariante por T y tiene dimensión mayor que dim W. Aplicando este proceso a $W = W_0$ se tiene β_1. Si $W_1 = W_0 + Z(\beta_1; T)$ es aún propio, entonces aplicando el proceso a W_1 se obtiene β_2. Se continúa de esta forma. Como dim $W_k >$ dim W_{k-1}, se debe alcanzar $W_r = V$ en no más de dim V etapas.

Parte 2. *Sean $\beta_1, \ldots, \beta_r$ vectores no nulos que satisfacen las condiciones* (*a*) *y* (*b*) *de la parte* 1. *Se fija k, $1 \leq k \leq r$. Sea β un vector cualquiera de V y sea $f = s(\beta; W_{k-1})$. Si*

$$f\beta = \beta_0 + \sum_{1 \leq i < k} g_i\beta_i, \qquad \beta_i \text{ en } W_i$$

entonces f divide a cada polinomio g_i y $\beta_0 = f\gamma_0$, donde γ_0 está en W_0.

Si $k = 1$, ello es justamente la afirmación de que W_0 es T-admisible. Para demostrar la afirmación para $k > 1$, se aplica el algoritmo de la división:

(7-4) $$g_i = fh_i + r_i, \qquad r_i = 0 \quad \text{o} \quad \operatorname{grd} r_i < \operatorname{grd} f.$$

Queremos mostrar que $r_i = 0$ para todo i. Sea

(7-5) $$\gamma = \beta - \sum_{1}^{k-1} h_i\beta_i.$$

Como $\gamma - \beta$ está en W_k,

$$s(\gamma; W_{k-1}) = s(\beta; W_{k-1}) = f.$$

Además.

(7-6) $$f\gamma = \beta_0 + \sum_{1}^{k-1} r_i\beta_i.$$

Supóngase que algún r_i es distinto de 0. Se llegará a una contradicción. Sea j el mayor índice i para el que $r_i \neq 0$. Entonces

(7-7) $$f\gamma = \beta_0 + \sum_{1}^{j} r_i\beta_i, \qquad r_j \neq 0 \quad \text{y} \quad \text{grd } r_j < \text{grd } f.$$

Sea $p = s(\gamma; W_{j-1})$. Como W_{k-1} contiene a W_{j-1}, el conductor $f = s(\gamma; W_{k-1})$ debe dividir a p:

$$p = fg.$$

Aplicando $g(T)$ a ambos miembros de (7-7):

(7-8) $$p\gamma = gf\gamma = gr_j\beta_j + g\beta_0 + \sum_{1 \leq i < j} gr_i\beta_i.$$

Por definición, $p\gamma$ está en W_{j-1}, y los dos últimos términos del segundo miembro de (7-8) están en W_{j-1}. Por tanto, $gr_j\beta_j$ está en W_{j-1}. Ahora, por la condición (b) de la parte 1:

$$\begin{aligned} \text{grd }(gr_j) &\geq \text{grd } s(\beta_j; W_{j-1}) \\ &= \text{grd } p_j \\ &\geq \text{grd } s(\gamma; W_{j-1}) \\ &= \text{grd } p \\ &= \text{grd }(fg). \end{aligned}$$

Con lo que grd $r_j >$ grd f, y ello contradice la elección de j. Se sabe ahora que f divide a cada g_i y luego que $\beta_0 = f\gamma$. Como W_0 es T-admisible, $\beta_0 = f\gamma_0$, donde γ_0 está en W_0. De paso notamos que la parte 2 es una reafirmación de la observación de que cada uno de los subespacios $W_1, W_2, \ldots, W_r$ es T-admisible.

Parte 3. *Existen vectores no nulos* $\alpha_1, \ldots, \alpha_r$ *en* V *que satisfacen las condiciones* (i) *y* (ii) *del Teorema* 3.

Se comienza con los vectores $\beta_1, \ldots, \beta_r$, como en la parte 1. Fijo k, $1 \leq k \leq r$, se aplica la parte 2 al vector $\beta = \beta_k$ y al T-conductor $f = p_k$. Obtenemos

(7-9) $$p_k\beta_k = p_k\gamma_0 + \sum_{1 \leq i < k} p_kh_i\beta_i$$

donde γ_0 está en W_0 y los $h_1, \ldots, h_{k-1}$ son polinomios. Sea

(7-10) $$\alpha_k = \beta_k - \gamma_0 - \sum_{1 \leq i < k} h_i\beta_i.$$

Como $\beta_k - \alpha_k$ está en W_{k-1},

(7-11) $$s(\alpha_k; W_{k-1}) = s(\beta_k; W_{k-1}) = p_k$$

y como $p_k\alpha_k = 0$, se tiene

(7-12) $$W_{k-1} \cap Z(\alpha_k; T) = \{0\}.$$

Como cada α_k satisface (7-11) y (7-12), se sigue que

$$W_k = W_0 \oplus Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_k; T)$$

y que p_k es el T-anulador de α_k. En otras palabras, los vectores $\alpha_1, \ldots, \alpha_r$ definen la misma sucesión de subespacios $W_1, W_2, \ldots$ que los vectores $\beta_1, \ldots, \beta_r$ y los T-conductores $p_k = s(\alpha_k; W_{k-1})$ tienen las mismas propiedades maximales [condición (b) de la parte 1]. Los vectores $\alpha_1, \ldots, \alpha_r$ tienen además la propiedad de que los subespacios W_0, $Z(\alpha_1; T)$, $Z(\alpha_2; T)$, ... son independientes. Es, por tanto, fácil verificar la condición (ii) en el Teorema 3. Como $p_i\alpha_i = 0$ para todo i, se tiene la relación trivial

$$p_k\alpha_k = 0 + p_1\alpha_1 + \cdots + p_{k-1}\alpha_{k-1}.$$

Aplicando la parte 2 con los $\beta_1, \ldots, \beta_k$ remplazados por los $\alpha_1, \ldots, \alpha_k$ y con $\beta = \alpha_k$ se tiene como conclusión que p_k divide a cada p_i, con $i < k$.

Parte 4. El número r y los polinomios $p_1, \ldots, p_r$ están unívocamente determinados por las condiciones del Teorema 3.

Supóngase que además de los vectores $\alpha_1, \ldots, \alpha_r$ en el Teorema 3, se tengan vectores no nulos $\gamma_1, \ldots, \gamma_s$ con los respectivos T-anuladores $g_1, \ldots, g_s$, tales que

(7-13) $$\begin{aligned} V &= W_0 \oplus Z(\gamma_1; T) \oplus \cdots \oplus Z(\gamma_s; T) \\ g_k &\text{ divide } g_{k-1}, \qquad k = 2, \ldots, s. \end{aligned}$$

Se demostrará que $r = s$ y que $p_i = g_i$ para todo i.

Es fácil ver que $p_1 = g_1$. El polinomio g_1 está determinado por (7-13) como el T-conductor de V en W_0. Sea $S(V; W_0)$ la colección de polinomios f tales que $f\beta$ esté en W_0 para cada β en V; es decir, los polinomios f tales que la imagen de $f(T)$ esté contenida en W_0. Entonces $S(V; W_0)$ es un ideal no nulo en el álgebra de los polinomios. El polinomio g_1 es, por esta razón, el generador mónico de ese ideal. Todo β de V tiene la forma

$$\beta = \beta_0 + f_1\gamma_1 + \cdots + f_s\gamma_s$$

y así,

$$g_1\beta = g_1\beta_0 + \sum_1^s g_1f_i\gamma_i.$$

Como cada g_i divide a g_1, se tiene que $g_1\gamma_i = 0$ para todo i y $g_1\beta = g_1\beta_0$ está en W_0. Con lo que g_1 está en $S(V; W_0)$. Como g_1 es el polinomio mónico de menor grado que aplica γ_1 en W_0, se ve que g_1 es el polinomio mónico de menor grado en el ideal $S(V; W_0)$. Por el mismo razonamiento, p_1 es el generador de tal ideal, con lo que $p_1 = g_1$.

Si f es un polinomio y W es un subespacio de V, usaremos la expresión abreviada fW para el conjunto de todos los vectores $f\alpha$ con α en W. Se deja para los ejercicios la demostración de los siguientes tres hechos:

1. $fZ(\alpha; T) = Z(f\alpha; T)$.
2. Si $V = V_1 \oplus \cdots \oplus V_k$, donde cada V_i es invariante por T, entonces $fV = fV_1 \oplus \cdots \oplus fV_k$.
3. Si α y γ tienen el mismo T-anulador, entonces $f\alpha$ y $f\gamma$ tienen el mismo T-anulador y (por tanto)

$$\dim Z(f\alpha; T) = \dim Z(f\gamma; T).$$

Ahora se procede por inducción para ver que $r = s$ y $p_i = g_i$ para $i = 2, \ldots, r$. El razonamiento consiste en contar las dimensiones en la forma correcta. Se dará la demostración de que si $r \geq 2$, entonces $p_2 = g_2$, y con ello la inducción será clara. Supóngase que $r \geq 2$. Entonces

$$\dim W_0 + \dim Z(\alpha_1; T) < \dim V.$$

Como se sabe que $p_1 = g_1$, se sabe que $Z(\alpha_1; T)$ y $Z(\gamma_1; T)$ tienen la misma dimensión. Por tanto,

$$\dim W_0 + \dim Z(\gamma_1; T) < \dim V$$

lo que muestra que $s \geq 2$. Ahora tiene sentido preguntarse si cabe o no $p_2 = g_2$. De las dos descomposiciones de V se obtienen dos descomposiciones del subespacio p_2V:

$$(7\text{-}14) \qquad \begin{aligned} p_2V &= p_2W_0 \oplus Z(p_2\alpha_1; T) \\ p_2V &= p_2W_0 \oplus Z(p_2\gamma_1; T) \oplus \cdots \oplus Z(p_2\gamma_s; T). \end{aligned}$$

Hemos usado los hechos (1) y (2) anteriores y el hecho de que $p_2\alpha_i = 0$, $i \geq 2$. Como se sabe que $p_1 = g_1$, (3), anteriormente mencionado, dice que $Z(p_2\alpha_1; T)$ y $Z(p_2\gamma_1, T)$ tienen la misma dimensión. Luego se desprende de (7-14) que

$$\dim Z(p_2\gamma_i; T) = 0, \qquad i \geq 2.$$

Concluimos que $p_2\gamma_2 = 0$ y que g_2 divide a p_2. El razonamiento puede invertirse para demostrar que p_2 divide a g_2. Por tanto, $p_2 = g_2$. ∎

Corolario. *Si T es un operador lineal sobre un espacio vectorial de dimensión finita, entonces todo subespacio T-admisible tiene un subespacio complementario que es también invariante por T.*

Demostración. Sea W_0 un subespacio admisible de V. Si $W_0 = V$, el complemento que se busca es $\{0\}$. Si W_0 es propio, se aplica el Teorema 3 y se hace

$$W_0' = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T).$$

Entonces W_0' es invariante por T y $V = W_0 \oplus W_0'$. ∎

Corolario. *Sea T un operador lineal sobre un espacio vectorial V de dimensión finita.*

(a) *Existe un vector* α *en* V *tal que el* T*-anulador de* α *es el polinomio minimal de* T.

(b) T *tiene un vector cíclico si, y solo si, los polinomios característico y minimal de* T *son idénticos.*

Demostración. Si $V = \{0\}$, los resultados son trivialmente verdaderos. Si $V \neq \{0\}$, sea

$$V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T) \tag{7-15}$$

donde los T-anuladores $p_1, \ldots, p_r$ son tales que p_{k-1} divide a p_k, $1 \leq k \leq r - 1$. Como observamos en la demostración del Teorema 3, se sigue fácilmente que p_1 es el polinomio minimal de T, es decir, el T-conductor de V en $\{0\}$. Con ello se ha demostrado (a).

Vimos en la Sección 7.1 que si T tiene un vector cíclico el polinomio minimal de T coincide con el polinomio característico. El contenido de (b) está en el recíproco. Se elige un α cualquiera, como en (a). Si el grado del polinomio minimal es dim V, entonces $V = Z(\alpha; T)$. ∎

Teorema 4 (teorema de Cayley-Hamilton generalizado). *Sea* T *un operador lineal sobre un espacio vectorial* V *de dimensión finita. Sean* p *y* f *los polinomios minimal y característico de* T*, respectivamente.*

(i) p *divide a* f;
(ii) p *y* f *tienen los mismos factores primos, salvo multiplicidades.*
(iii) *Si*

$$p = f_1^{r_1} \cdots f_k^{r_k} \tag{7-16}$$

es la factorización prima de p*, entonces*

$$f = f_1^{d_1} \cdots f_k^{d_k} \tag{7-17}$$

donde d_i *es la nulidad de* $f(T)^{r_i}$ *dividida por el grado de* f_i.

Demostración. No se considera el caso trivial $V = \{0\}$. Para demostrar (i) y (ii) se considera una descomposición cíclica (7-15) de V, que se obtiene del Teorema 3. Como observamos en la demostración del segundo corolario, $p_1 = p$. Sea U_i la restricción de T a $Z(\alpha_i; T)$. Entonces U_i tiene un vector cíclico, y así p_i es el polinomio minimal y el polinomio característico de U_i. Por tanto, el polinomio característico f es el producto $f = p_1, \ldots, p_r$. Esto es evidente por la forma bloque de (6-14) que la matriz de T toma en una base apropiada. Evidentemente, $p_1 = p$ divide a f, y ello demuestra (i). Es claro que cualquier divisor primo de p es divisor primo de f. Recíprocamente, un divisor primo de $f = p_1 \cdots p_r$ debe dividir a uno de los factores p_i que a su vez divide a p_1.

Sea (7-16) la factorización prima de p. Se emplea el teorema de descomposición prima (Teorema 12 del Capítulo 6) que dice que, si V_i es el espacio nulo de $f_i(T)^{r_i}$, entonces

$$V = V_1 \oplus \cdots \oplus V_k \tag{7-18}$$

y $f_i^{r_i}$ es el polinomio minimal del operador T_i, obtenido por restricción de T al subespacio (invariante) V_i. Se aplica la parte (ii) del presente teorema al operador T_i. Como su polinomio minimal es una potencia del primo f_i, el polinomio característico de T_i tiene la forma $f_i^{d_i}$, donde $d_i \geq r_i$. Evidentemente

$$d_i = \frac{\dim V_i}{\text{grd } f_i}$$

y (casi por definición) $\dim V_i$ = nulidad $f_i(T)^{r_i}$. Como T es la suma directa de los operadores $T_1, \ldots, T_k$, el polinomio característico f es el producto

$$f = f_1^{d_1} \cdots f_k^{d_k}. \quad \blacksquare$$

Corolario. *Si T es un operador lineal nilpotente sobre un espacio vectorial de dimensión n, entonces el polinomio característico de T es x^n.*

Se desea ver ahora el análogo del teorema de descomposición cíclica para matrices. Si se tiene el operador T y la descomposición en suma directa del Teorema 3, sea $\mathcal{B}_i$ la «base ordenada cíclica»

$$\{\alpha_i, T\alpha_i, \ldots, T^{k_i-1}\alpha_i\}$$

para $Z(\alpha_i; T)$. Aquí k_i representa la dimensión de $Z(\alpha_i; T)$, esto es, el grado del anulador p_i. La matriz del operador inducido T_i en la base ordenada $\mathcal{B}_i$ es la matriz asociada del polinomio p_i. Así, si se hace que $\mathcal{B}$ sea la base ordenada de V que es la unión de las $\mathcal{B}_i$ dispuestas en el orden $\mathcal{B}_1, \ldots, \mathcal{B}_r$, entonces la matriz de T en la base ordenada $\mathcal{B}$ será

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_r \end{bmatrix} \tag{7-19}$$

donde A_i es la $k_i \times k_i$ matriz asociada de p_i. De una matriz $n \times n$, A, que es la suma directa (7-19) de matrices asociadas de polinomios mónicos no escalares $p_1, \ldots, p_r$, tales que p_{i+1} divide a p_i para $i = 1, \ldots, r-1$, se dice que está en **forma racional.** El teorema de descomposición cíclica dice lo siguiente respecto a las matrices.

Teorema 5. *Sea F un cuerpo y sea B una matriz $n \times n$ sobre F. Entonces B es semejante sobre el cuerpo F a una, y solo a una, matriz que está en forma racional.*

Demostración. Sea T el operador lineal sobre F^n representado por B en la base ordenada canónica. Como acabamos de observar, existe una base ordenada de F^n en que T está representado por una matriz A en forma racional. Entonces B es semejante a esta matriz A. Supóngase que B es semejante sobre F a otra matriz C que está en la forma racional. Esto quiere decir simplemente que existe una cierta base ordenada de F^n en la que el operador T está representado por la matriz C. Si C es la suma directa de matrices asociadas C_i de

polinomios mónicos $g_1, \ldots, g_s$, tal que g_{i+1} divide a g_i para $i = 1, \ldots, s-1$, entonces es evidente que se tienen vectores $\beta_1, \ldots, \beta_s$, no nulos, en V con T-anuladores $g_1, \ldots, g_s$ tales que

$$V = Z(\beta_1; T) \oplus \cdots \oplus Z(\beta_s; T).$$

Pero por la unicidad del teorema de descomposición cíclica, los polinomios g_i son idénticos a los polinomios p_i que definen la matriz A. Así, $C = A$. ∎

Los polinomios $p_1, \ldots, p_r$ son llamados los **factores invariantes** de la matriz B. En la Sección 7.4 se describirá un algoritmo para calcular los factores invariantes de una matriz dada B. El que sea posible calcular estos polinomios por medio de un número finito de operaciones racionales en los elementos de B, es la razón por la cual la forma racional recibe ese nombre.

Ejemplo 2. Supóngase que V es un espacio vectorial bidimensional sobre el cuerpo F, y T sea un operador lineal sobre V. Las posibilidades de descomposición cíclica en subespacios para T son muy limitadas. En efecto, si el polinomio minimal de T es de grado 2, es igual al polinomio característico de T, y T tiene un vector cíclico. Así que existe cierta base ordenada de V en la que T está representado por la matriz asociada de su polinomio característico. Si, por otro lado, el polinomio minimal de T es de grado 1, T es un múltiplo escalar del operador identidad. Si $T = cI$, entonces para dos vectores linealmente independientes cualesquiera, α_1 y α_2, en V, se tiene

$$V = Z(\alpha_1; T) \oplus Z(\alpha_2; T)$$
$$p_1 = p_2 = x - c.$$

Para matrices, este análisis dice que toda matriz 2×2 sobre el cuerpo F es semejante sobre F exactamente a una matriz de los tipos

$$\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}, \quad \begin{bmatrix} 0 & -c_0 \\ 1 & -c_1 \end{bmatrix}.$$

Ejemplo 3. Sea T el operador lineal sobre R^3 representado por la matriz

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

en la base ordenada canonica. Se ha calculado antes que el polinomio característico de T es $f = (x-1)(x-2)^2$, y que el polinomio minimal de T es $p = (x-1)(x-2)$. Así se sabe que en la descomposición cíclica de T el primer vector α_1 tendrá a p como T-anulador. Como se ha operado en un espacio tridimensional, puede haber solo otro vector, α_2, el cual debe generar un subespacio cíclico de dimensión 1, es decir, debe ser un vector propio de T. Su T-anulador p_2 debe ser $(x-2)$, ya que debemos tener que $pp_2 = f$. Obsérvese que esto dice inmediatamente que la matriz A es semejante a la matriz

$$B = \begin{bmatrix} 0 & -2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

esto es, que T está representado por B en alguna base ordenada. ¿Cómo se pueden encontrar los vectores adecuados α_1 y α_2? Bien, sabemos que cualquier vector que genera un subespacio T-cíclico de dimensión 2 es un vector adecuado α_1. De modo que sea ϵ_1. Tenemos

$$T\epsilon_1 = (5, -1, 3)$$

que no es un múltiplo escalar de ϵ_1; luego $Z(\epsilon_1; T)$ tiene dimensión 2. Este espacio consta de todos los vectores $a\epsilon_1 + b(T\epsilon_1)$:

$$a(1, 0, 0) + b(5, -1, 3) = (a + 5b, -b, 3b)$$

o todos los vectores (x_1, x_2, x_3) que satisfacen $x_3 = -3x_2$. Ahora lo que se desea es un vector α_2 tal que $T\alpha_2 = 2\alpha_2$ y $Z(\alpha_2; T)$ es disjunto de $Z(\alpha_1; T)$. Como α_2 es un vector característico para T, el espacio $Z(\alpha_2; T)$ será simplemente el espacio unidimensional generado por α_2, y así, lo que se requiere es que α_2 no esté en $Z(\alpha_1; T)$. Si $\alpha = (x_1, x_2, x_3)$, se puede fácilmente calcular que $T\alpha = 2\alpha$ si, y solo si, $x_1 = 2x_2 + 2x_3$. Así, $\alpha_2 = (2, 1, 0)$ satisface $T\alpha_2 = 2\alpha_2$ y genera un subespacio T-cíclico disjunto de $Z(\alpha_1; T)$. El lector podrá verificar directamente que la matriz de T en la base ordenada

$$\{(1, 0, 0), (5, -1, 3), (2, 1, 0)\}$$

es la matriz B anteriormente mencionada.

Ejemplo 4. Supóngase que T es un operador lineal diagonalizable sobre V. Es interesante relacionar una descomposición cíclica de T con una base que diagonaliza la matriz de T. Sean $c_1, \ldots, c_k$ los valores propios distintos de T, y sea V_i el espacio de los vectores propios asociados con el valor característico c_i. Entonces

$$V = V_1 \oplus \cdots \oplus V_k$$

y si $d_i = \dim V_i$, entonces

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

es el polinomio característico de T. Si α es un vector de V, es fácil relacionar el subespacio cíclico $Z(\alpha; T)$ con los subespacios $V_1, \ldots, V_k$. Existen vectores $\beta_1, \ldots, \beta_k$ únicos, tales que β_i está en V_i, y

$$\alpha = \beta_1 + \cdots + \beta_k.$$

Como $T\beta_i = c_i\beta_i$, se tiene

$$f(T)\alpha = f(c_1)\beta_1 + \cdots + f(c_k)\beta_k \tag{7-20}$$

para todo polinomio f. Dados escalares arbitrarios $t_1, \ldots, t_k$, existe un polinomio f tal que $f(c_i) = t_i$, $1 \leq i \leq k$. Por tanto. $Z(\alpha; T)$ es justamente el subespacio generado por los vectores $\beta_1, \ldots, \beta_k$. ¿Cuál es el anulador de α? De acuerdo con (7-20) se tiene $f(T)\alpha = 0$ si, y solo si, $f(c_i)\beta_i = 0$ para cada i. En otras palabras, $f(T)\alpha = 0$ siempre que $f(c_i) = 0$ para cada i tal que $\beta_i \neq 0$. En consecuencia, el anulador de α es el producto

(7-21) $$\prod_{\beta_i \neq 0} (x - c_i).$$

Ahora, sean $\mathfrak{B}_i = \{\beta_1^i, \ldots, \beta_{d_i}^i\}$ una base ordenada de V_i. Sea

$$r = \max_i d_i.$$

Se definen los vectores $\alpha_1, \ldots, \alpha_r$ por

(7-22) $$\alpha_j = \sum_{d_i \geq j} \beta_j^i, \qquad 1 \leq j \leq r.$$

El subespacio cíclico $Z(\alpha_j; T)$ es el subespacio generado por los vectores β_j^i, cuando i describe los índices para los cuales $d_i \geq j$. El T-anulador de α_j es

(7-23) $$p_j = \prod_{d_i \geq j} (x - c_i).$$

Se tiene

$$V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T)$$

pues cada β_j^i pertenece a uno, y solo a uno, de los subespacios $Z(\alpha_1; T), \ldots, Z(\alpha_r; T)$, y $\mathfrak{B} = (\mathfrak{B}_1, \ldots, \mathfrak{B}_k)$ es una base ordenada de V. Por (7-23), p_{j+1} divide a p_j.

Ejercicios

1. Sea T el operador lineal sobre F^2 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Sea $\alpha_1 = (0, 1)$. Demostrar que $F^2 \neq Z(\alpha_1; T)$ y que existe un vector α_2, no nulo, en F^2 con $Z(\alpha_2; T)$ disjunto de $Z(\alpha_1; T)$.

2. Sea T un operador lineal sobre el espacio V de dimensión finita y sea R la imagen por T.

(a) Demostrar que R tiene un subespacio T-invariante complementario si, y solo si, R es independiente del espacio nulo N de T.

(b) Si R y N son independientes, demostrar que N es el subespacio único T-invariante complementario de R.

3. Sea T el operador lineal sobre R^3 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Sea W el espacio nulo de $T - 2I$. Demostrar que W no tiene subespacio T-invariante complementario. (*Sugerencia*: Sea $\beta = \epsilon_1$ y obsérvese que $(T - 2I)\beta$ está en W. Demostrar que no existe α en W con $(T - 2I)\beta = (T - 2I)\alpha$.)

4. Sea T el operador lineal sobre F^4 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} c & 0 & 0 & 0 \\ 1 & c & 0 & 0 \\ 0 & 1 & c & 0 \\ 0 & 0 & 1 & c \end{bmatrix}.$$

Sea W el espacio nulo de $T - cI$.

(a) Demostrar que W es el subespacio generado por ϵ_4.

(b) Hallar el generador mónico de los ideales $S(\alpha_4; W)$, $S(\alpha_3; W)$, $S(\alpha_2; W)$ y $S(\alpha_1; W)$.

5. Sea T un operador lineal sobre el espacio vectorial sobre el cuerpo F. Si f es un polinomio sobre F y α está en V, sea $f\alpha = f(T)\alpha$. Si $V_1, \ldots, V_k$ son subespacios invariantes bajo T y $V = V_1 \oplus \cdots \oplus V_k$, demostrar que

$$fV = fV_1 \oplus \cdots \oplus fV_k.$$

6. Sean T, V y F como en el Ejercicio 5. Supóngase que α y β son vectores en V que tienen el mismo T-anulador. Demostrar que, para cualquier polinomio f, los vectores $f\alpha$ y $f\beta$ tienen el mismo T-anulador.

7. Hallar los polinomios minimales y las formas racionales de cada una de las siguientes matrices reales

$$\begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} c & 0 & -1 \\ 0 & c & 1 \\ -1 & 1 & c \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

8. Sea T el operador lineal sobre R^3 representado en la base ordenada canónica por

$$\begin{bmatrix} 3 & -4 & -4 \\ -1 & 3 & 2 \\ 2 & -4 & -3 \end{bmatrix}.$$

Hallar los vectores no nulos $\alpha_1, \ldots, \alpha_r$ que satisfacen las condiciones del Teorema 3.

9. Sea A la matriz real

$$A = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 1 & 3 \\ -3 & -3 & -5 \end{bmatrix}$$

Hallar una matriz real 0, inversible, tal que $P^{-1}AP$ esté en la forma racional

10. Sea F un subcuerpo de los números complejos y sea T el operador lineal sobre F^4 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & a & 2 & 0 \\ 0 & 0 & b & 2 \end{bmatrix}.$$

Hallar el polinomio característico de T. Considerar los casos $a = b = 1$, $a = b = 0$. $a = 0$, $b = 1$. En cada uno de los casos, hallar el polinomio minimal de T y los vectores $\alpha_1, \ldots, a_r$, no nulos que satisfacen las condiciones del Teorema 3.

11. Demostrar que si A y B son matrices 3×3 sobre el cuerpo F, una condición necesaria y suficiente para que A y B sean semejantes sobre F es que tengan el mismo polinomio

(7-21) $$\prod_{\beta_i \neq 0} (x - c_i).$$

Ahora, sean $\mathcal{B}_i = \{\beta_1^i, \ldots, \beta_{d_i}^i\}$ una base ordenada de V_i. Sea

$$r = \max_i d_i.$$

Se definen los vectores $\alpha_1, \ldots, \alpha_r$ por

(7-22) $$\alpha_j = \sum_{d_i \geq j} \beta_j^i, \qquad 1 \leq j \leq r.$$

El subespacio cíclico $Z(\alpha_j; T)$ es el subespacio generado por los vectores β_j^i, cuando i describe los índices para los cuales $d_i \geq j$. El T-anulador de α_j es

(7-23) $$p_j = \prod_{d_i \geq j} (x - c_i).$$

Se tiene

$$V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T)$$

pues cada β_j^i pertenece a uno, y solo a uno, de los subespacios $Z(\alpha_1; T), \ldots, Z(\alpha_r; T)$, y $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ es una base ordenada de V. Por (7-23), p_{j+1} divide a p_j.

Ejercicios

1. Sea T el operador lineal sobre F^2 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Sea $\alpha_1 = (0, 1)$. Demostrar que $F^2 \neq Z(\alpha_1; T)$ y que existe un vector α_2, no nulo, en F^2 con $Z(\alpha_2; T)$ disjunto de $Z(\alpha_1; T)$.

2. Sea T un operador lineal sobre el espacio V de dimensión finita y sea R la imagen por T.

(a) Demostrar que R tiene un subespacio T-invariante complementario si, y solo si, R es independiente del espacio nulo N de T.

(b) Si R y N son independientes, demostrar que N es el subespacio único T-invariante complementario de R.

3. Sea T el operador lineal sobre R^3 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Sea W el espacio nulo de $T - 2I$. Demostrar que W no tiene subespacio T-invariante complementario. (*Sugerencia*: Sea $\beta = \epsilon_1$ y obsérvese que $(T - 2I)\beta$ está en W. Demostrar que no existe α en W con $(T - 2I)\beta = (T - 2I)\alpha$.)

4. Sea T el operador lineal sobre F^4 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} c & 0 & 0 & 0 \\ 1 & c & 0 & 0 \\ 0 & 1 & c & 0 \\ 0 & 0 & 1 & c \end{bmatrix}.$$

Sea W el espacio nulo de $T - cI$.

(a) Demostrar que W es el subespacio generado por ϵ_4.

(b) Hallar el generador mónico de los ideales $S(\alpha_4; W)$, $S(\alpha_3; W)$, $S(\alpha_2; W)$ y $S(\alpha_1; W)$.

5. Sea T un operador lineal sobre el espacio vectorial sobre el cuerpo F. Si f es un polinomio sobre F y α está en V, sea $f\alpha = f(T)\alpha$. Si $V_1, \ldots, V_k$ son subespacios invariantes bajo T y $V = V_1 \oplus \cdots \oplus V_k$, demostrar que

$$fV = fV_1 \oplus \cdots \oplus fV_k.$$

6. Sean T, V y F como en el Ejercicio 5. Supóngase que α y β son vectores en V que tienen el mismo T-anulador. Demostrar que, para cualquier polinomio f, los vectores $f\alpha$ y $f\beta$ tienen el mismo T-anulador.

7. Hallar los polinomios minimales y las formas racionales de cada una de las siguientes matrices reales

$$\begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} c & 0 & -1 \\ 0 & c & 1 \\ -1 & 1 & c \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

8. Sea T el operador lineal sobre R^3 representado en la base ordenada canónica por

$$\begin{bmatrix} 3 & -4 & -4 \\ -1 & 3 & 2 \\ 2 & -4 & -3 \end{bmatrix}.$$

Hallar los vectores no nulos $\alpha_1, \ldots, \alpha_r$ que satisfacen las condiciones del Teorema 3.

9. Sea A la matriz real

$$A = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 1 & 3 \\ -3 & -3 & -5 \end{bmatrix}$$

Hallar una matriz real 0, inversible, tal que $P^{-1}AP$ esté en la forma racional

10. Sea F un subcuerpo de los números complejos y sea T el operador lineal sobre F^4 representado en la base ordenada canónica por la matriz

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & a & 2 & 0 \\ 0 & 0 & b & 2 \end{bmatrix}.$$

Hallar el polinomio característico de T. Considerar los casos $a = b = 1$, $a = b = 0$, $a = 0$, $b = 1$. En cada uno de los casos, hallar el polinomio minimal de T y los vectores $\alpha_1, \ldots, a_r$, no nulos que satisfacen las condiciones del Teorema 3.

11. Demostrar que si A y B son matrices 3×3 sobre el cuerpo F, una condición necesaria y suficiente para que A y B sean semejantes sobre F es que tengan el mismo polinomio

característico y el mismo polinomio minimal. Dar un ejemplo que muestre que esto es falso para matrices 4×4.

12. Sea F un subcuerpo del cuerpo de los números complejos y sean A y B matrices $n \times n$ sobre F. Demostrar que si A y B son semejantes sobre el cuerpo de los números complejos, son semejantes sobre F. (*Sugerencia*: Demostrar que la forma racional de A es la misma si se considera A como matriz sobre F o como matriz sobre C; lo mismo para B).

13. Sea A una matriz $n \times n$ de elementos complejos. Demostrar que si todo valor propio de A es real, A es semejante a una matriz de elementos reales.

14. Sea T un operador lineal sobre el espacio V de dimensión finita. Demostrar que existe un vector α en V con esta propiedad: Si f es un polinomio y $f(T)\alpha = 0$, entonces $f(T) = 0$. (Tal vector α se llama un **vector separador** para el álgebra de los polinomios en T.) Cuando T tiene un vector cíclico, dar una demostración directa de que cualquier vector cíclico es un vector separador para el álgebra de los polinomios en T.

15. Sea F un subcuerpo del cuerpo de los números complejos y sea A una matriz $n \times n$ sobre F. Sea p el polinomio minimal para A. Si se considera a A como matriz sobre C, entonces A tiene un polinomio minimal f como matriz $n \times n$ sobre C. Usar un teorema sobre ecuaciones lineales para demostrar que $p = f$. ¿Se puede ver también cómo se deduce esto del teorema de descomposición cíclica?

16. Sea A una matriz $n \times n$ con elementos reales tal que $A^2 + I = 0$. Demostrar que n es par y, si $n = 2k$, entonces A es semejante sobre el cuerpo de los números reales a una matriz de forma bloque

$$\begin{bmatrix} 0 & I \\ I- & 0 \end{bmatrix}$$

donde I es la matriz unidad $k \times k$.

17. Sea T un operador lineal sobre un espacio vectorial V de dimensión finita. Supóngase que

(a) el polinomio minimal de T es una potencia de un polinomio irreducible;

(b) el polinomio minimal es igual al polinomio característico.

Demostrar que ningún subespacio no trivial invariante por T tiene un subespacio complementario invariante por T.

18. Si T es un operador lineal diagonalizable, entonces cada subespacio invariante por T tiene un subespacio complementario invariante por T.

19. Sea T un operador lineal sobre el espacio V de dimensión finita. Demostrar que T tiene un vector cíclico si, y solo si, es cierto lo siguiente: cada operador lineal U que conmuta con T es un polinomio en T.

20. Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea T un operador lineal en V. Se pregunta cuándo se verifica que cada vector no nulo de V es un vector cíclico para T. Demostrar que este es el caso si, y solo si, el polinomio característico para T es irreducible sobre F.

21. Sea A una matriz $n \times n$ con elementos *reales*. Sea T el operador lineal sobre R^n representado por A en la base ordenada canónica y sea U el operador lineal sobre C^n representado por A en la base ordenada canónica. Usar el resultado del Ejercicio 20 para demostrar lo siguiente: si los únicos subespacios invariantes por T son R^n y el subespacio cero, entonces U es diagonalizable.

7.3. La forma de Jordan

Supóngase que N es un operador lineal nilpotente en el espacio V de dimensión finita. Se considera la descomposición cíclica para N que se obtiene del Teorema 3. Se tiene un entero positivo r y r vectores no negativos $\alpha_1, \ldots, \alpha_r$ en V con N-anuladores $p_1, \ldots, p_r$, tales que

$$V = Z(\alpha_1; N) \oplus \cdots \oplus Z(\alpha_r; N)$$

y p_{i+1} divide a p_i para $i = 1, \ldots, r - 1$. Como N es nilpotente, el polinomio minimal es x^k para algún $k \leq n$. Así, todo p_i es de la forma $p_i = x^k$ y la condición de divisibilidad simplemente dice que

$$k_1 \geq k_2 \geq \cdots \geq k_r.$$

Por supuesto, $k_1 = k$ y $k_r \geq 1$. La matriz asociada de x^{k_i} es la matriz $k_i \times k_i$.

$$A_i = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \tag{7-24}$$

Así, el Teorema 3 da una base ordenada para V, en la cual la matriz de N es suma directa de matrices elementales nilpotentes (7-24), cuyas dimensiones decrecen cuando i crece. Por lo que se ve que, asociados con una matriz nilpotente $n \times n$, hay un entero positivo r y r enteros positivos $k_1, \ldots, k_r$, de modo que $k_1 + \cdots + k_r = n$ y $k_i \geq k_{i+1}$, y estos enteros positivos determinan la forma racional de la matriz, es decir, determinan la matriz, salvo semejanza.

He aquí un punto que deseamos destacar respecto al operador nilpotente. El entero positivo r es justamente la nulidad de N; en efecto, el espacio nulo tiene como base los r vectores

$$N^{k_i - 1}\alpha_i. \tag{7-25}$$

Si α está en el espacio nulo de N, se puede escribir α en la forma

$$\alpha = f_1\alpha_1 + \cdots + f_r\alpha_r$$

donde f_i es un polinomio, cuyo grado se puede suponer menor que k_i. Como $N\alpha = 0$, para todo i tenemos

$$\begin{aligned} 0 &= N(f_i\alpha_i) \\ &= Nf_i(N)\alpha_i \\ &= (xf_i)\alpha_i. \end{aligned}$$

Así, xf_i es divisible por x^k, y como grd $(f_i) < k_i$, ello quiere decir que

$$f_i = c_i x^{k_i - 1}$$

donde c_i es cierto escalar. Pero entonces

$$\alpha = c_1(x^{k_1 - 1}\alpha_1) + \cdots + c_r(x^{k_r - 1}\alpha_r)$$

que muestra que los vectores (7-25) forman una base para el espacio nulo de N. El lector puede observar que este hecho es también claro desde el punto de vista matricial.

Lo que se desea hacer ahora es combinar lo establecido para operadores o matrices nilpotentes con el teorema de descomposición prima del Capítulo 6. La situación es la siguiente: supóngase que T es un operador lineal sobre V y que el polinomio característico de T se puede factorizar sobre F como sigue

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

donde los $c_1, \ldots, c_k$ son elementos distintos de F y $d_i \geq 1$. Entonces el polinomio minimal de T será

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}$$

donde $1 \leq r_i \leq d_i$. Si W_i es el espacio nulo de $(T - c_i I)^{r_i}$, entonces el teorema de descomposición prima dice que

$$V = W_1 \oplus \cdots \oplus W_k$$

y que el operador T_i inducido en W_i por T tiene polinomio minimal $(x - c_i)^{r_i}$. Sea N_i el operador lineal en W_i definido por $N_i = T_i - c_i I$. Entonces N_i es nilpotente y tiene un polinomio minimal x^{r_i}. En W_i, T actúa como N_i más el escalar c_i veces el operador identidad. Supóngase que elegimos una base del subespacio W_i que corresponda a la descomposición cíclica del operador nilpotente N_i. Entonces la matriz de T_i en esa base ordenada será la suma directa de matrices

$$\begin{bmatrix} c & 0 & \cdots & 0 & 0 \\ 1 & c & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ & & & c & \\ 0 & 0 & \cdots & 1 & c \end{bmatrix} \tag{7-26}$$

cada cual con $c = c_i$. Además, las dimensiones de estas matrices decrecerán al ir de izquierda a derecha. Una matriz de la forma (7-26) se llama una **matriz elemental de Jordan con valor propio** c. Ahora, si se consideran en conjunto todas las bases para W_i, se obtiene una base ordenada de V. Se describirá la matriz A de T en esta base ordenada.

La matriz A es la suma directa

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix} \tag{7-27}$$

de matrices $A_1, \ldots, A_k$. Toda A_i es de la forma

$$A_i = \begin{bmatrix} J_1^{(i)} & 0 & \cdots & 0 \\ 0 & J_2^{(i)} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & J_{n_i}^{(i)} \end{bmatrix}$$

donde cada $J_j^{(i)}$ es la matriz elemental de Jordan con valor propio c_i. También en cada A_i la dimensión de las matrices $J_j^{(i)}$ decrece cuando j crece. Se dirá que una matriz $n \times n$, A, que satisface todas las condiciones descritas hasta ahora (para escalares distintos $c_1, \ldots, c_k$) está en **forma de Jordan.**

Hemos observado que si T es un operador lineal cuyo polinomio característico se puede factorizar completamente sobre el cuerpo escalar, entonces existe una base ordenada de V, en la cual T está en forma de Jordan. Deseamos mostrar ahora que esta matriz es algo unívocamente asociado con T, salvo el orden en que se escriben los valores propios de T. Es decir, si dos matrices están en forma de Jordan y son semejantes, entonces difieren solo en la ordenación de los escalares c_i.

La unicidad se ve como sigue. Supóngase que existe una base ordenada de V en la cual T está representado por la matriz de Jordan A, descrita anteriormente. Si A_i es una matriz $d_i \times d_i$, entonces d_i es claramente la multiplicidad de c_i como raíz de polinomio característico de A o de T. En otras palabras, el polinomio característico de T es

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}.$$

Esto demuestra que $c_1, \ldots, c_k$ y $d_1, \ldots, d_k$ son únicos, salvo el orden en que se escriben. El que A sea la suma directa de las matrices A_i, nos da una descomposición en suma directa $V = W_1 \oplus \cdots \oplus W_k$ invariante por T. Obsérvese ahora que W_i debe ser el espacio nulo de $(T - c_iI)^n$, donde $n = \dim V$; en efecto, $A_i - c_iI$ es evidentemente nilpotente y $A_j - c_iI$ es no singular para $j \neq i$. Así vemos que los subespacios W_i son únicos. Si T_i es el operador inducido en W_i por T, entonces la matriz A_i está unívocamente determinada como la forma racional de $(T_i - c_iI)$.

Queremos ahora hacer algunas observaciones más respecto al operador T y la matriz de Jordan A que representa a T en una base ordenada. Ellas son:

(1) Cada elemento de A, que no esté en la diagonal principal o inmediatamente debajo de ésta es 0. En la diagonal de A están los k valores propios distintos $c_1, \ldots, c_k$ de T. También c_i está repetido d_i veces, donde d_i es la multiplicidad de c_i como raíz del polinomio característico, es decir, $d_i = \dim W_i$.

(2) Para todo i, la matriz A_i es la suma directa de n_i matrices elementales de Jordan $J_j^{(i)}$ con valor propio c_i. El número n_i es precisamente la dimensión del espacio de vectores propios asociados al valor propio c_i. En efecto, n_i es el número de bloques elementales nilpotentes en la forma racional de $(T_i - c_iI)$, y es así igual a la dimensión del espacio nulo de $(T - c_iI)$. En particular, obsérvese que T es diagonalizable si, y solo si, $n_i = d_i$ para todo i.

(3) Para todo i el primer bloque $J_1^{(i)}$ en la matriz A_i es una matriz $r_i \times r_i$, donde r_i es la multiplicidad de c_i como raíz del polinomio *minimal* de T. Esto se desprende de que el polinomio minimal del operador nilpotente $(T_i - c_iI)$ es x^{r_i}.

Por supuesto, como de costumbre, se tienen resultados semejantes para las matrices. Si B es una matriz $n \times n$ sobre el cuerpo F y si el polinomio característico de B se factoriza completamente sobre F, entonces B es semejante sobre

F a una matriz $n \times n$, A, en la forma de Jordan, y A es única salvo el orden de los valores propios. Se dice que A es la **forma de Jordan** de B.

Obsérvese también que si F es un cuerpo algebraicamente cerrado, entonces las acotaciones anteriores son válidas para todo operador lineal sobre un espacio vectorial de dimensión finita sobre F, o para toda matriz $n \times n$ sobre F. Así, por ejemplo, toda matriz $n \times n$ sobre el cuerpo de los números complejos es semejante a una matriz, esencialmente única, en forma de Jordan.

Ejemplo 5. Supóngase que T es un operador lineal sobre C^2. El polinomio característico para T es $(x - c_1)(x - c_2)$, si los números complejos c_1, c_2 son distintos, o $(x - c_1)^2$. En el primer caso T es diagonalizable y está representado en una base ordenada por

$$\begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}.$$

En el segundo caso el polinomio minimal de T puede ser $(x - c)$, y entonces $T = cI$, o puede ser $(x - c)^2$, en cuyo caso T está representado en una base ordenada por la matriz

$$\begin{bmatrix} c & 0 \\ 1 & c \end{bmatrix}.$$

Así toda matriz 2×2 sobre el cuerpo de los complejos es semejante a una matriz de alguno de los dos tipos anteriores, posiblemente con $c_1 = c_2$.

Ejemplo 6. Sea A una matriz compleja 3×3:

$$A = \begin{bmatrix} 2 & 0 & 0 \\ a & 2 & 0 \\ b & c & -1 \end{bmatrix}.$$

El polinomio característico de A es evidentemente $(x - 2)^2(x + 1)$. O este es el polinomio minimal, en cuyo caso A es semejante a

$$\begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

o el polinomio minimal es $(x - 2)(x + 1)$, en cuyo caso A es semejante a

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Ahora

$$(A - 2I)(A + I) = \begin{bmatrix} 0 & 0 & 0 \\ 3a & 0 & 0 \\ ac & 0 & 0 \end{bmatrix}$$

y así A es semejante a una matriz diagonal si, y solo si, $a = 0$.

Ejemplo 7. Sea

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & a & 2 \end{bmatrix}.$$

El polinomio característico de A es $(x - 2)^4$. Como A es suma directa de dos matrices 2×2, es claro que el polinomio minimal de A es $(x - 2)^2$. Ahora si $a = 0$ o si $a = 1$, entonces la matriz A está en forma de Jordan. Obsérvese que las dos matrices que se obtienen para $a = 0$ y $a = 1$ tienen el mismo polinomio característico, pero no son semejantes. Y no lo son porque para la primera matriz el espacio solución de $(A - 2I)$ tiene dimensión 3, mientras que para la segunda matriz tiene dimensión 2.

Ejemplo 8 Las ecuaciones diferenciales lineales con coeficientes constantes (Ejemplo 14, Capítulo 6) dan una bonita ilustración de la forma de Jordan. Sean $a_0, \ldots, a_{n-1}$ números complejos y sea V el espacio de todas las funciones n veces derivables en un intervalo del eje real que satisfacen la ecuación diferencial

$$\frac{d^n f}{dx^n} + a_{n-1}\frac{d^{n-1}f}{dx^{n-1}} + \cdots + a_1\frac{df}{dx} + a_0 f = 0.$$

Sea D el operador derivación. Entonces V es invariante por D, ya que V es el espacio nulo de $p(D)$, donde

$$p = x^n + \cdots + a_1 x + a_0.$$

¿Cuál es la forma de Jordan para el operador derivación sobre V?

Sean $c_1, \ldots, c_k$ las raíces complejas distintas de p:

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}.$$

Sea V_i el espacio nulo de $(D - c_i I)^{r_i}$, esto es, el conjunto de las soluciones de la ecuación diferencial

$$(D - c_i I)^{r_i} f = 0.$$

Entonces, como se observó en el Ejemplo 15, Capítulo 6, el teorema de descomposición prima dice que

$$V = V_1 \oplus \cdots \oplus V_k.$$

Sea N_i la restricción de $D - c_i I$ a V_i. La forma de Jordan para el operador D (en V) está entonces determinada por las formas racionales de los operadores nilpotentes $N_1, \ldots, N_k$ en los espacios $V_1, \ldots, V_k$.

Así, lo que se debe conocer (para los distintos valores de c) es la forma racional del operador $N = (D - cI)$ en el espacio V_c, que consta de las soluciones de la ecuación

$$(D - cI)^r f = 0.$$

¿Cuántos bloques elementales nilpotentes habrá en la forma racional para N? Su número será la nulidad de N; es decir, la dimensión del espacio propio asociado al valor propio c. Esa dimensión es 1, pues cualquier función que satisface la ecuación diferencial

$$Df = cf$$

es un múltiplo escalar de la función exponencial $h(x) = e^{cx}$. Por tanto, el operador N (sobre el espacio V_c) tiene un vector cíclico. Una buena elección para un vector cíclico es $g = x^{r-1}h$:

$$g(x) = x^{r-1}e^{cx}.$$

Esto da

$$\begin{aligned} Ng &= (r-1)x^{r-2}h \\ &\vdots \\ N^{r-1}g &= (r-1)!h \end{aligned}$$

Lo visto en el párrafo anterior muestra que la forma de Jordan para D (en el espacio V) es suma directa de k matrices elementales de Jordan, una por cada raíz c_i.

Ejercicios

1. Sean N_1 y N_2 matrices 3×3 nilpotentes sobre el cuerpo F. Demostrar que N_1 y N_2 son semejantes si, y solo si, tienen el mismo polinomio minimal.

2. Usar el resultado del Ejercicio 1 y la forma de Jordan para demostrar lo siguiente. Sean A y B matrices $n \times n$ sobre el cuerpo F que tienen el *mismo* polinomio característico

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

y el mismo polinomio minimal. Si ningún d_i es mayor que 3, entonces A y B son semejantes.

3. Si A es una matriz compleja 5×5 con polinomio característico

$$f = (x - 2)^3(x + 7)^2$$

y polinomio minimal $p = (x - 2)^2(x + 7)$, ¿cuál es la forma de Jordan para A?

4. ¿Cuántas posibles formas de Jordan hay para una matriz compleja 6×6 de polinomio característico $(x + 2)^4(x - 1)^2$?

5. El operador derivación sobre el espacio de los polinomios de grado menor o igual a 3 está representado en la base ordenada «natural» por la matriz

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

¿Cuál es la forma de Jordan para esta matriz? (F un subespacio de los números complejos.)

6. Sea A la matriz compleja

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 0 \\ 1 & 1 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

Hallar la forma de Jordan para A.

7. Si A es una matriz $n \times n$ sobre el cuerpo F con polinomio característico

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

¿cuál es la traza de A?

8. Clasificar, por semejanza, todas las matrices complejas 3×3, A, tales que $A^3 = I$.

9. Clasificar, por semejanza, todas las matrices complejas $n \times n$, A, tales que $A^n = I$.

10. Sea n un entero positivo, $n \geq 2$, y sea N una matriz $n \times n$ sobre el cuerpo F tal que $N^n = 0$, pero $N^{n-1} \neq 0$. Demostrar que N no tiene raíz cuadrada; es decir, que no existe una matriz $n \times n$, A, tal que $A^2 = N$.

11. Sean N_1 y N_2 matrices nilpotentes 6×6 sobre el cuerpo F. Supóngase que N_1 y N_2 tienen el mismo polinomio minimal y la misma nulidad. Demostrar que N_1 y N_2 son semejantes. Demostrar que esto no es cierto para matrices nilpotentes 7×7.

12. Usar el resultado del Ejercicio 11 y de la forma de Jordan para demostrar lo siguiente. Sean A y B matrices $n \times n$ sobre el cuerpo F que tienen el mismo polinomio característico

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

y el mismo polinomio minimal. Supóngase que también para todo i los espacios solución de $(A - c_iI)$ y $(B - c_iI)$ tienen la misma dimensión. Si ningún d_i es mayor que 6, entonces A y B son semejantes.

13. Si N es una matriz nilpotente $k \times k$ elemental, es decir, $N^k = 0$, pero $N^{k-1} \neq 0$, demostrar que N^t es semejante a N. Usar entonces la forma de Jordan para demostrar que toda matriz compleja $n \times n$ es semejante a su transpuesta.

14. ¿Cuál es el error en la siguiente demostración? Si A es una matriz compleja $n \times n$ tal que $A^t = -A$, entonces A es 0. (*Demostración*: Sea J la forma de Jordan de A. Como $A^t = -A$, entonces $J^t = -J$. Pero J es triangular, con lo que $J^t = -J$ implica que todo elemento J es cero. Como $J = 0$ y A es semejante a J, se desprende que $A = 0$.) (Dar un ejemplo de matriz A no nula tal que $A^t = -A$.)

15. Si N es una matriz nilpotente 3×3 sobre C, demostrar que $A = I + \frac{1}{2}N - \frac{1}{8}N^2$ satisface a $A^2 = I + N$, es decir, que A es la raíz cuadrada de $I + N$. Usar la serie binómica de $(1 + t)^{1/2}$ para obtener una fórmula parecida para una raíz cuadrada de $I + N$, donde N es una matriz nilpotente $n \times n$ sobre C.

16. Usar el resultado del Ejercicio 15 para demostrar que si c es un número complejo no nulo y N es una matriz compleja nilpotente, entonces $(cI + N)$ tiene una raíz cuadrada. En seguida usar la forma de Jordan para demostrar que toda matriz compleja $n \times n$, no singular, tiene una raíz cuadrada.

7.4. Cálculo de factores invariantes

Supóngase que A es una matriz $n \times n$ con elementos en el cuerpo F. Se desea encontrar un método para calcular los factores invariantes $p_1, \ldots, p_r$ que definen la forma racional para A. Se comienza con el caso sencillo en que A es la matriz adjunta (7.2) de un polinomio mónico

$$p = x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0.$$

En la Sección 7.1 se vio que p es tanto el polinomio minimal como el polinomio característico para la matriz adjunta A. Ahora queremos hacer un cálculo directo que muestre que p es el polinomio característico para A. En este caso

$$xI - A = \begin{bmatrix} x & 0 & 0 & \cdots & 0 & c_0 \\ -1 & x & 0 & \cdots & 0 & c_1 \\ 0 & -1 & x & \cdots & 0 & c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & x & c_{n-2} \\ 0 & 0 & 0 & \cdots & -1 & x + c_{n-1} \end{bmatrix}.$$

Se suma x veces la fila n a la fila $(n - 1)$, con lo que se quita x del lugar $(n - 1, n - 1)$ sin cambiar el valor del determinante. Después se suma x veces la nueva fila $n - 1$ a la fila $n - 2$. Se continúa sucesivamente este proceso hasta que todas las x de la diagonal principal hayan desaparecido. El resultado es la matriz

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & x^n + \cdots + c_1x + c_0 \\ -1 & 0 & 0 & \cdots & 0 & x^{n-1} + \cdots + c_2x + c_1 \\ 0 & -1 & 0 & \cdots & 0 & x^{n-2} + \cdots + c_3x + c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & x^2 + c_{n-1}x + c_{n-2} \\ 0 & 0 & 0 & \cdots & -1 & x + c_{n-1} \end{bmatrix}$$

que tiene el mismo determinante que $xI - A$. El elemento de la parte superior derecha de esta matriz es el polinomio p. Se puede simplificar la última columna sumando a ella múltiplos de las otras columnas:

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & p \\ -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 0 \end{bmatrix}$$

Se multiplica cada una de las primeras $(n - 1)$ columnas por -1 y se efectúan entonces $(n - 1)$ intercambios de columnas adyacentes con el objeto de llevar

la columna n a la primera posición. El efecto total de los $2n - 2$ cambios de signos deja inalterado el valor del determinante. Se obtiene la matriz

$$(7\text{-}28) \qquad \begin{bmatrix} p & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Es entonces claro que $p = \det(xI - A)$.

Vamos a mostrar que, para toda matriz $n \times n$, A, hay una sucesión de operaciones de filas y columnas que transforman $xI - A$ en una matriz como (7-28), en la que los factores invariantes de A aparecen en la diagonal principal. Se explicarán completamente las operaciones que se deben utilizar.

Se operará en $F[x]^{m \times n}$, colección de las matrices $n \times n$ con elementos que son polinomios sobre el cuerpo F. Si M es una de tales matrices, una **operación elemental de fila** sobre M es alguna de las siguientes:

1. Multiplicación de una fila de M por un escalar no nulo de F.
2. Remplazo de la r-ésima fila por la suma de la fila r más f veces la fila s, donde f es cualquier polinomio sobre F y $r \neq s$.
3. Intercambio de dos filas de M.

La operación inversa de una operación elemental de fila es una operación elemental de fila del mismo tipo. Obsérvese que no se puede hacer una afirmación tal si en (1) se permiten polinomios no escalares. Una **matriz elemental** $m \times m$, esto es, una matriz elemental de $F[x]^{m \times n}$ es aquella que se puede obtener de la matriz unidad $m \times m$ por medio de una operación elemental de fila simple. Es evidente que cada operación elemental de fila sobre M puede efectuarse multiplicando M a la izquierda por una matriz elemental $m \times m$ apropiada; en efecto, si e es la operación, entonces

$$e(M) = e(I)M.$$

Sean M, N matrices de $F[x]^{m \times n}$. Se dice que N es **equivalente por filas** a M si N puede obtenerse a partir de M por una sucesión finita de operaciones elementales de fila:

$$M = M_0 \rightarrow M_1 \rightarrow \cdots \rightarrow M_k = N.$$

Evidentemente, N es equivalente por filas a M si, y solo si, M es equivalente por filas a N, razón por la cual se usará la terminología «M y N son equivalentes por filas». Si N es equivalente por filas a M, entonces

$$N = PM$$

donde la matriz $m \times m$, P, es un producto de matrices elementales:

$$P = E_1 \cdots E_k.$$

En particular, P es una matriz inversible, con inversa

$$P^{-1} = E_k^{-1} \cdots E_1^{-1}.$$

Claro que la inversa de E_j proviene de la operación elemental de fila inversa.

Todo esto es precisamente como en el caso de las matrices con elementos en F. Se obtienen resultados elementales paralelos a los del Capítulo 1. Por tanto, el siguiente problema que se plantea es introducir una forma escalón reducida por filas para matrices polinomiales. Aquí se presenta un nuevo obstáculo. ¿Cómo se reduce por filas una matriz? La primera etapa es aislar el primer elemento de la fila 1 y dividir cada elemento de la fila 1 por ese elemento. Ello no se puede hacer (necesariamente) cuando la matriz tiene elementos polinomios. Como se verá en el siguiente teorema, se puede soslayar esta dificultad en algunos casos; sin embargo, no existe una forma reducida por filas completamente apropiada para la matriz general de $F[x]^{m\times n}$. Si se introducen también operaciones de columna y se estudia el tipo de equivalencia que resulta de permitir el uso de ambos tipos de operaciones, se puede obtener una forma canónica muy útil para toda matriz. El instrumento básico es el siguiente.

Lema. *Sea M una matriz de $F[x]^{m\times n}$ que tiene algunos elementos no nulos en su primera columna, y sea p el máximo común divisor de los elementos de la columna 1 de M. Entonces M es equivalente por filas a la matriz N que tiene*

$$\begin{bmatrix} p \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

como primera columna.

Demostración. Demostraremos algo más de lo que se ha enunciado. Mostraremos que existe un algoritmo para encontrar N, es decir, una receta que una máquina calculadora podría utilizar para calcular N en un número finito de etapas. Ante todo, se necesita cierta notación.

Sea M una matriz $m \times n$ con elementos en $F[x]$ que tiene una primera columna no nula

$$M_1 = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}.$$

Se define

$$l(M_1) = \min_{f \neq 0} \operatorname{grd} f_i \tag{7-29}$$

$$p(M_1) = \text{M.c.d. } (f_1, \ldots, f_m).$$

Sea j un índice tal que $\operatorname{grd} f_j = l(M_1)$. Para precisar, sea j el menor índice i para el que $\operatorname{grd} f_i = l(M_1)$. Para dividir cada f_i por f_j se hace

$$f_i = f_j g_i + r_i, \qquad r_i = 0 \quad \text{o} \quad \operatorname{grd} r_i < \operatorname{grd} f_j. \tag{7-30}$$

Para cada i diferente de j, se remplaza la fila i de M por la fila i menos g_i veces la fila f. Se multiplica la fila j por el inverso del coeficiente dominante de f_j, y

entonces se intercambian las filas j y 1. El resultado de estas operaciones es una matriz M' que tiene como primera columna

$$(7\text{-}31) \qquad M_1' = \begin{bmatrix} \tilde{f}_j \\ r_2 \\ \vdots \\ r_{j-1} \\ r_1 \\ r_{j+1} \\ \vdots \\ r_m \end{bmatrix}.$$

donde $\tilde{f}_j$ es el polinomio mónico que se obtiene por normalización de f_j para que tenga coeficiente dominante 1. Hemos dado un procedimiento bien definido para asociar a cada matriz M una matriz M' con las siguientes propiedades:

(a) M' es equivalente por filas a M.
(b) $p(M_1') = p(M_1)$.
(c) O, $l(M_1') < l(M_1)$ o

$$M_1' = \begin{bmatrix} p(M_1) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Es fácil comprobar (b) y (c) de (7-30) y (7-31). La propiedad (c) es justamente otro modo de establecer que, o existe un i tal que $r_i \neq 0$ y grd $r_i <$ grd f_j, o bien $r_i = 0$ para todo i y $\tilde{f}_j$ es (por tanto) el máximo común divisor de $f_1, \ldots, f_m$.

La demostración del lema es ahora muy simple. Se comienza con la matriz M aplicándole el procedimiento anterior para obtener M'. La propiedad (c) dice que M' servirá como la matriz N en el lema, o, $l(M_1') < l(M_1)$. En el último caso se aplica el procedimiento a M' para obtener la matriz $M^{(2)} = (M')'$. Si $M^{(2)}$ no es el N conveniente, se forma $M^{(3)} = (M^{(2)})'$ y así sucesivamente. El caso es que las desigualdades estrictas

$$l(M_1) > l(M_1') > l(M_1^{(2)}) > \cdots$$

no pueden seguir mucho. Después de no más que $l(M_1)$ iteraciones de este procedimiento, se debe llegar a una matriz $M^{(k)}$ que tiene las propiedades que se buscan. ▮

Teorema 6. *Sea P una matriz $m \times m$ con elementos en el álgebra de polinomios $F[x]$. Las siguientes afirmaciones son equivalentes:*

(i) *P es inversible.*
(ii) *El determinante de P es un polinomio escalar no nulo.*

(iii) *P es equivalente por filas a la matriz unidad $m \times m$.*
(iv) *P es un producto de matrices elementales.*

Demostración. Ciertamente (i) implica (ii), ya que la función determinante es multiplicativa y los únicos polinomios inversibles de $F[x]$ son los polinomios escalares no nulos. En realidad, en el Capítulo 5 se usó el adjunto para hacer ver que (i) y (ii) son equivalentes. Este razonamiento da una demostración diferente de que (i) se desprende de (ii). Se completará el ciclo

$$\begin{array}{ccc} (i) & \rightarrow & (ii) \\ \uparrow & & \downarrow \\ (iv) & \leftarrow & (iii). \end{array}$$

La única implicación que no es obvia es que (iii) se desprende de (ii).

Supóngase (ii) y considérese la primera columna de P. Ella contiene ciertos polinomios $p_1, \ldots, p_m$, y

$$\text{M.c.d. } (p_1, \ldots, p_m) = 1$$

ya que cada común divisor de $p_1, \ldots, p_m$ debe dividir (al escalar) det P. Aplicando el lema anterior a P se tiene la matriz

$$Q = \begin{bmatrix} 1 & a_2 & \cdots & a_m \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{bmatrix} \tag{7-32}$$

que es equivalente por filas a P. Una operación elemental de fila altera el determinante de una matriz por (a lo más) un factor escalar no nulo. Así, det Q es un polinomio escalar no nulo. Evidentemente, la matriz $(m-1) \times (m-1)$, B, en (7-32) tiene el mismo determinante que Q. Por tanto, se aplica el lema anterior a B. Si se continúa de esta manera por m etapas, se obtiene la matriz triangular superior

$$R = \begin{bmatrix} 1 & a_2 & \cdots & a_m \\ 0 & 1 & \cdots & b_m \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

que es equivalente por filas a R. Obviamente, R es equivalente por filas a la matriz unidad $m \times m$. ▮

Corolario. *Sean M y N matrices $m \times n$ con elementos en el álgebra de polinomios $F[x]$. Entonces N es equivalente por filas a M si, y solo si,*

$$N = PM$$

donde P es una matriz inversible $m \times m$ con elementos en $F[x]$.

Definimos ahora **operaciones elementales de columna y equivalencia por columnas** de un modo análogo a las operaciones de fila y equivalencia por filas. No se necesitan nuevos tipos de matrices elementales, ya que la clase de matri-

…e que se puede obtener al hacer una operación elemental por columna en la matriz unidad es la misma que la clase obtenida al usar una sola operación elemental por fila.

Definición. *La matriz N es* **equivalente** *a la matriz M si se puede pasar de M a N por medio de una sucesión de operaciones*

$$M = M_0 \rightarrow M_1 \rightarrow \cdots \rightarrow M_k = N$$

cada una de las cuales es una operación elemental de fila o una operación elemental de columna.

Teorema 7. *Sean M y N matrices $m \times n$ con elementos en el álgebra de polinomios $F[x]$. Entonces N es equivalente a M si, y solo si,*

$$N = PMQ$$

donde P es una matriz inversible de $F[x]^{m \times m}$ y Q es una matriz inversible de $F[x]^{n \times n}$.

Teorema 8. *Sea A una matriz $n \times n$ con elementos en el cuerpo F y sean $p_1, \ldots, p_r$ los factores invariantes de A. La matriz $xI - A$ es equivalente a la matriz diagonal $n \times n$ cuyos elementos en la diagonal son $p_1, \ldots, p_r, 1, 1, \ldots, 1$.*

Demostración. Existe una matriz inversible $n \times n$, P, con elementos en F tal que PAP^{-1} está en la forma racional, esto es, tiene la forma bloque

$$PAP^{-1} = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{bmatrix}$$

donde A_i es la matriz asociada al polinomio p_i. De acuerdo con el Teorema 7, la matriz

(7-33) $$P(xI - A)P^{-1} = xI - PAP^{-1}$$

es equivalente a $xI - A$. Ahora

(7-34) $$xI - PAP^{-1} = \begin{bmatrix} xI - A_1 & 0 & \cdots & 0 \\ 0 & xI - A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & xI - A_r \end{bmatrix}$$

donde los varios I representan las distintas matrices unidad de dimensión apropiada. Al comienzo de esta sección se discutió que $xI - A_i$ es equivalente a la matriz

$$\begin{bmatrix} p_i & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

De (7-33) y (7-34) es claro que $xI - A$ es equivalente a una matriz diagonal que tiene los polinomios p_i y $(n - r)$ unos en la diagonal principal. Por una sucesión de intercambios de filas y columnas se pueden ordenar los elementos de aquella diagonal en el orden que se desee; por ejemplo, $p_1, \ldots, p_r, 1, \ldots, 1$. ▌

El Teorema 8 no da un medio eficaz para calcular los divisores elementales $p_1, \ldots, p_r$, ya que la demostración depende del teorema de descomposición cíclica. Daremos ahora un algoritmo explícito para reducir una matriz de polinomios a forma diagonal. El Teorema 8 sugiere que se pueden disponer también esos elementos sucesivos en la diagonal principal de modo que uno divida al otro.

Definición. *Sea N una matriz de $F[x]^{m \times n}$. Se dice que N está en la* **forma normal** *(de Smith) si*

(a) *todo elemento fuera de la diagonal principal de N es* 0;

(b) *en la diagonal principal de N aparecen (en orden) los polinomios $f_1, \ldots, f_l$ de modo tal que f_k divide a f_{k+1}, $1 \leq k \leq l - 1$.*

En la definición el número l es $l = \min(m, n)$. Los elementos de la diagonal principal son $f_k = N_{kk}$, $k = 1, \ldots, l$.

Teorema 9. *Sea M una matriz $m \times n$ con elementos en el álgebra de polinomios $F[x]$. Entonces M es equivalente a una matriz N que está en la forma normal.*

Demostración. Si $M = 0$, no hay nada que demostrar. Si $M \neq 0$ se dará un algoritmo para encontrar una matriz M' que sea equivalente a M y que tiene la forma

$$(7\text{-}35) \qquad M' = \begin{bmatrix} f_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & R & \\ 0 & & & \end{bmatrix}$$

donde R es una matriz $(m - 1) \times (n - 1)$ y f_1 divide cada elemento de R. Con ello se habrá terminado la demostración, ya que se puede aplicar el mismo procedimiento a R para obtener f_2, etc.

Sea $l(M)$ el mínimo de los grados de los elementos no nulos de M. Hállese la primera columna que contenga un elemento de grado $l(M)$ e intercambiar esa columna con la columna 1. Se llamará a la matriz resultante $M^{(0)}$. Describimos un procedimiento para hallar una matriz de la forma

$$(7\text{-}36) \qquad \begin{bmatrix} g & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & S & \\ 0 & & & \end{bmatrix}$$

que es equivalente a $M^{(0)}$. Empezamos por aplicar a la matriz $M^{(0)}$ el procedimiento del lema anterior al Teorema 6, procedimiento que se llamará PL6. Resulta una matriz

$$M^{(1)} = \begin{bmatrix} p & a & \cdots & b \\ 0 & c & \cdots & d \\ \vdots & \vdots & & \vdots \\ 0 & e & \cdots & f \end{bmatrix}. \tag{7-37}$$

Si los elementos $a, \ldots, b$ son todos nulos, bien. Si no, se usa el análogo de PL6 para la primera fila, procedimiento que se llamará PL6'. El resultado es una matriz

$$M^{(2)} = \begin{bmatrix} q & 0 & \cdots & 0 \\ a' & c' & \cdots & e' \\ \vdots & \vdots & & \vdots \\ b' & d' & \cdots & f' \end{bmatrix} \tag{7-38}$$

donde q es el máximo común divisor de p, a, ..., b. Para obtener $M^{(2)}$ puede haberse alterado o no la buena forma de la columna 1. Si se ha alterado, se aplica PL6 otra vez. Y esto es lo que se busca. En no más de $l(M)$ etapas

$$M^{(0)} \xrightarrow{\text{PL6}} M^{(1)} \xrightarrow{\text{PL6}'} M^{(2)} \xrightarrow{\text{PL6}} \cdots \longrightarrow M^{(t)}$$

se debe llegar a una matriz $M^{(t)}$ que tiene la forma (7-36), ya que en cada etapa sucesiva se tiene que $l(M^{(k+1)}) < l(M^{(k)})$. Se designa el proceso que se acaba de definir por P7-36:

$$M^{(0)} \xrightarrow{\text{P7-36}} M^{(t)}.$$

En (7-36), el polinomio g puede o no dividir cada elemento de S. Si no, hallar la primera columna que tiene un elemento que no sea divisible por g y sumar tal columna a la columna 1. La primera nueva columna contiene a g y un elemento $gh + r$ donde $r \neq 0$ y $\operatorname{grd} r < \operatorname{grd} g$. Aplicando el proceso P7-36, el resultado será otra matriz de la forma (7-36), donde el grado del correspondiente g ha decrecido.

Ahora es obvio que en un número finito de etapas se obtendrá (7-35), es decir, se llegará a una matriz de la forma (7-36) donde el grado de g no puede reducirse más. ∎

Queremos hacer ver que la forma normal asociada a una matriz M es única. Se vieron dos cosas que dan la clave de cómo los polinomios $f_1, \ldots, f_l$ en el Teorema 9 están determinados unívocamente por M. Primera, las operaciones elementales de fila y de columna no cambian el valor del determinante de una matriz cuadrada en más que un factor escalar no nulo. Segunda, las operaciones elementales de fila y de columna no cambian el máximo común divisor de cada elemento de una matriz.

Definición. *Sea M una matriz $m \times n$ con elementos en $F[x]$. Si $1 \leq k \leq$ min (m, n), se define $\delta_k(M)$ como el máximo común divisor de los determinantes de todas las submatrices $k \times k$ de M.*

Recuérdese que una submatriz $k \times k$ de M es aquella que se obtiene suprimiendo $m - k$ filas y $n - k$ columnas de M. En otras palabras, se eligen ciertos k-tuples

$$\begin{aligned} I &= (i_1, \ldots, i_k), \qquad 1 \leq i_1 < \cdots < i_k \leq m \\ J &= (j_1, \ldots, j_k), \qquad 1 \leq j_1 < \cdots < j_k \leq n \end{aligned}$$

y se observa la matriz formada por esas filas y columnas de M. Estamos interesados en los determinantes

$$D_{I,J}(M) = \det \begin{bmatrix} M_{i_1j_1} & \cdots & M_{i_1j_k} \\ \vdots & & \vdots \\ M_{i_kj_1} & \cdots & M_{i_kj_k} \end{bmatrix}. \tag{7-39}$$

El polinomio $\delta_k(M)$ es el máximo común divisor de los polinomios $D_{I,J}(M)$, cuando I y J describen todos los posibles k-tuples.

Teorema 10. *Si M y N son matrices $m \times n$ equivalentes, con elementos en $F[x]$, entonces*

$$\delta_k(M) = \delta_k(N), \qquad 1 \leq k \leq \min(m, n). \tag{7-40}$$

Demostración. Es suficiente demostrar que una sola operación elemental de fila e no cambia a δ_k. Como la inversa de e es también una operación elemental de fila, basta con demostrar lo siguiente: Si un polinomio f divide a todo $D_{I,J}(M)$, entonces f divide a $D_{I,J}(e(M))$ para todos los k-tuples I y J.

Como estamos considerando una operación de fila, sean $\alpha_1, \ldots, \alpha_m$ las filas de M y usemos la notación

$$D_J(\alpha_{i_1}, \ldots, \alpha_{i_k}) = D_{I,J}(M).$$

Dados I y J, ¿cuál es la relación entre $D_{I,J}(M)$ y $D_{I,J}(e(M))$? Se consideran los tres tipos de operaciones e:

(a) multiplicación de la fila r por un escalar c no nulo;
(b) remplazo de la fila r por la fila r más g veces la fila s, $r \neq s$;
(c) intercambio de las filas r y s, $r \neq s$.

Se deja a un lado por el momento el tipo (c) y se tratará de los tipos (a) y (b) que solo cambian la fila r. Si r no es uno de los índices $i_1, \ldots, i_k$, entonces

$$D_{I,J}(e(M)) = D_{I,J}(M).$$

Si r está entre los índices $i_1, \ldots, i_k$, entonces para los dos casos tenemos

$$\begin{aligned} \text{(a)}\quad D_{I,J}(e(M)) &= D_J(\alpha_{i_1}, \ldots, c\alpha_r, \ldots, \alpha_{i_k}) \\ &= cD_J(\alpha_{i_1}, \ldots, \alpha_r, \ldots, \alpha_{i_k}) \\ &= cD_{I,J}(M); \end{aligned}$$

$$\begin{aligned} \text{(b)}\quad D_{I,J}(e(M)) &= D_J(\alpha_{i_1}, \ldots, \alpha_r + g\alpha_s, \ldots, \alpha_{i_k}) \\ &= D_{I,J}(M) + gD_J(\alpha_{i_1}, \ldots, \alpha_s, \ldots, \alpha_{i_k}). \end{aligned}$$

Para las operaciones del tipo (a) es claro que cualquier f que divide a $D_{I,J}(M)$ también divide a $D_{I,J}(e(M))$. Para el caso de la operación del tipo (b) obsérvese que

$$D_J(\alpha_{i_1}, \ldots, \alpha_s, \ldots, \alpha_{i_k}) = 0, \quad \text{si } s = i_j \text{ para algún } j$$
$$D_J(\alpha_{i_1}, \ldots, \alpha_s, \ldots, \alpha_{i_k}) = \pm D_{I',J}(M), \quad \text{si } s \neq i_j \text{ para todo } j.$$

La I' en la última ecuación es el k-tuple $(i_1, \ldots, s, \ldots, i_k)$ ordenado en orden creciente. Ahora es evidente que, si f divide cada $D_{I,J}(M)$, entonces f divide a cada $D_{I,J}(e(M))$.

Las operaciones del tipo (c) pueden tratarse más o menos por el mismo razonamiento dado, o usando el hecho de que tal operación puede ser efectuada por una sucesión de operaciones de los tipos (a) y (b). ▮

Corolario. *Toda matriz M de $F[x]^{m\times n}$ es equivalente precisamente a una matriz N que está en forma normal. Los polinomios $f_1, \ldots, f_l$, que aparecen en la diagonal principal de N son*

$$f_k = \frac{\delta_k(M)}{\delta_{k-1}(M)}, \qquad 1 \leq k \leq min\ (m, n)$$

donde, por conveniencia, se define $\delta_0(M) = 1$.

Demostración. Si N está en forma normal con elementos en la diagonal $f_1, \ldots, f_l$, es muy fácil ver que

$$\delta_k(N) = f_1 f_2 \cdots f_k. \quad ▮$$

Por supuesto que la matriz N en este último corolario se llama la **forma normal** de M. Los polinomios $f_1, \ldots, f_l$ son llamados a menudo los **factores invariantes** de M.

Supóngase que A es una matriz $n \times n$ con elementos en F, y sean $p_1, \ldots, p_r$ los factores invariantes de A. Se ve ahora que la forma normal de la matriz $xI - A$ tiene en la diagonal los elementos $1, 1, \ldots, 1, p_r, \ldots, p_1$. El último corolario dice qué son $p_1, \ldots, p_r$, en términos de las submatrices de $xI - A$. El número $n - r$ es el mayor k tal que $\delta_k(xI - A) = 1$. El polinomio minimal p_1 es el polinomio característico de A dividido por el máximo común divisor de los determinantes de todas las submatrices $(n - 1) \times (n - 1)$ de $xI - A$, etc.

Ejercicios

1. ¿Es verdadero o falso que toda matriz de $F[x]^{n\times n}$ es equivalente por filas a una matriz triangular superior?

2. Sea T un operador lineal sobre un espacio vectorial de dimensión finita y sea A la matriz de T en una base ordenada. Entonces T tiene un vector cíclico si, y solo si, los determinantes de las submatrices $(n - 1) \times (n - 1)$ de $xI - A$ son primos relativos.

3. Sea A una matriz $n \times n$ con elementos en el cuerpo F y sean $f_1, \ldots, f_n$ los elementos de la diagonal de la forma normal de $xI - A$. ¿Para qué matrices A es $f_1 \neq 1$?

4. Construir un operador lineal T con polinomio minimal $x^2(x-1)^2$ y polinomio característico $x^3(x-1)^4$. Describir la descomposición prima del espacio vectorial por T y hallar las proyecciones sobre los componentes primos. Hallar una base en la cual la matriz de T esté en forma de Jordan. Hallar también una descomposición explícita en suma directa del espacio subespacios T-cíclicos, como en el Teorema 3 y dar los factores invariantes.

5. Sea T el operador lineal sobre R^8 representado en la base canónica por la matriz

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & -1 & -1 & -1 & -1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

(a) Hallar el polinomio característico y los factores invariantes.

(b) Hallar la descomposición prima de R^8 por T y las proyecciones en los componentes primos. Hallar la descomposición cíclica de cada componente primo, como en el Teorema 3.

(c) Hallar la forma de Jordan de A.

(d) Hallar una descomposición en suma directa de R^8 en subespacios T-cíclicos, como en el Teorema 3. (*Sugerencia*: Un modo de hacerlo es usar el resultado en (b) y una apropiada generalización de las ideas estudiadas en el Ejemplo 4.)

7.5. *Resumen: operadores semisimples*

En los dos últimos capítulos nos hemos ocupado de un operador lineal simple T sobre un espacio vectorial V de dimensión finita. El programa ha consistido en descomponer T en suma directa de operadores lineales de naturaleza elemental, con el objeto de tener información detallada respecto a cómo «opera» T sobre el espacio V. Resumimos lo que se ha hecho hasta ahora.

Comenzamos estudiando T por medio de los valores propios y los vectores propios. Introdujimos operadores diagonalizables, operadores que pueden ser completamente descritos en términos de los valores y vectores propios. Observamos entonces que T podía no tener un vector propio simple. Aun en el caso de un cuerpo escalar algebraicamente cerrado, cuando todo operador lineal tiene al menos un vector propio, se observó que los vectores propios de T no generaban necesariamente el espacio.

Se demostró luego el teorema de descomposición cíclica, que expresa cualquier operador lineal como suma directa de operadores con un vector cíclico, sin hipótesis alguna acerca del cuerpo escalar. Si U es un operador lineal con un vector cíclico, existe una base $\{\alpha_1, \ldots, \alpha_n\}$ con

$$U\alpha_j = \alpha_{j+1}, \qquad j = 1, \ldots, n-1$$
$$U\alpha_n = -c_0\alpha_1 - c_1\alpha_2 - \cdots - c_{n-1}\alpha_n.$$

El efecto de U sobre esta base es entonces trasladar cada α_j al siguiente vector α_{j+1}, con excepción de que $U\alpha_n$ es una combinación lineal preestablecida de los vectores de la base. Como el operador lineal general T es suma directa de un número finito de tales operadores U, se obtuvo una descripción explícita y razonablemente elemental del efecto de T.

A continuación se aplicó el teorema de descomposición cíclica a operadores nilpotentes. Para el caso de un cuerpo escalar algebraicamente cerrado, se combinó esto con el teorema de descomposición prima para obtener la forma de Jordan. La forma de Jordan da una base $\{\alpha_1, \ldots, \alpha_n\}$ del espacio V tal que, para todo j, o bien $T\alpha_j$ es un múltiplo escalar de α_j, o bien $T\alpha_j = c\alpha_j + a_{j+1}$. Tal base describe ciertamente el efecto de T de modo explícito y elemental.

La importancia de la forma racional (o de la forma de Jordan) proviene de que existe, más bien que del hecho que se pueda calcular en casos determinados. Claro está que si se tiene un operador lineal específico T y se puede calcular su forma cíclica o de Jordan, es lo que hay que hacer, pues, teniendo esa forma, se puede deducir una gran cantidad de información acerca de T. Dos tipos diferentes de dificultades surgen en el cálculo de esas formas canónicas. Una es desde luego lo largo de los cálculos. Otra, que puede no haber un método para hacer los cálculos, incluso si se tienen el tiempo y la paciencia necesarios. La segunda dificultad surge, por ejemplo, al tratar de hallar la forma de Jordan de una matriz compleja. Sencillamente no hay un método bien definido para factorizar el polinomio característico, y así uno queda detenido al empezar. La forma racional no presenta esta dificultad. Como se vio en la Sección 7.4, hay un método bien definido para hallar la forma racional de una matriz $n \times n$ dada; con todo, los cálculos son por lo general extremadamente largos.

En el resumen de los resultados de estos dos últimos capítulos no se ha mencionado aún uno de los teoremas que se demostraron. Es el teorema que dice que si T es un operador lineal sobre un espacio vectorial de dimensión finita sobre un cuerpo algebraicamente cerrado, T se expresa unívocamente como suma de un operador diagonalizable y de un operador nilpotente que conmutan. Esto se demostró con el teorema de descomposición prima y con cierta información sobre los operadores diagonalizables. No es éste un teorema tan profundo como el de la descomposición cíclica o el de la existencia de la forma de Jordan, pero sí tiene aplicaciones importantes y útiles en algunas partes de la matemática. Para concluir este capítulo demostraremos un teorema análogo sin suponer que el cuerpo escalar es algebraicamente cerrado. Comenzaremos definiendo los operadores que desempeñan el papel de los operadores diagonalizables.

Definición. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea T un operador lineal sobre V. Se dice que T es* **semisimple** *si todo subespacio T-invariante tiene un subespacio complementario T-invariante.*

Lo que se va a demostrar es que, con ciertas restricciones sobre el cuerpo F, todo operador lineal T se expresa unívocamente en la forma $T = S + N$, don-

de S es semisimple, N es nilpotente y $SN = NS$. Primero se han de caracterizar los operadores semisimples por medio de sus polinomios minimales, y esta caracterización hará ver que, cuando F es algebraicamente cerrado, un operador es semisimple si, y solo si, es diagonalizable.

Lema. *Sea T un operador lineal sobre el espacio vectorial de dimensión finita V y sea $V = W_1 \oplus \cdots \oplus W_k$ la descomposición prima de T; es decir, si p es el polinomio minimal de T y $p = p_1^{r_1} \cdots p_k^{r_k}$ es la factorización prima de p, entonces W_j es el espacio nulo de $p_j(T)^{r_j}$. Sea W el subespacio de V que es invariante por T. Entonces*

$$W = (W \cap W_1) \oplus \cdots \oplus (W \cap W_k)$$

Demostración. Para la demostración se necesita recordar un corolario de la demostración del teorema de descomposición prima, en la Sección 6.8. Si $E_1, \ldots, E_k$ son las proyecciones asociadas a la descomposición $V = W_1 \oplus \cdots \oplus W_k$, entonces cada E_j es un polinomio en T. Esto es, existen polinomios $h_1, \ldots, h_k$ tales que $E_j = h_j(T)$.

Sea ahora W el subespacio que es invariante por T. Si α es un vector cualquiera de W, entonces $\alpha = \alpha_1 + \cdots + \alpha_k$, con α_j en W_j. Ahora $\alpha_j = E_j\alpha = h_j(T)\alpha$, y como W es invariante bajo T, cada α_j está también en W. Así cada vector α de W es de la forma $\alpha = \alpha_1 + \cdots + \alpha_k$, con α_j en la intersección $W \cap W_j$. Esta expresión es única, ya que $V = W_1 \oplus \cdots \oplus W_k$. Por tanto,

$$W = (W \cap W_1) \oplus \cdots \oplus (W \cap W_k). \quad \blacksquare$$

Lema. *Sea T un operador lineal sobre V y supóngase que el polinomio minimal de T es irreducible sobre el cuerpo escalar F. Entonces T es semisimple.*

Demostración. Sea W un subespacio de V invariante por T. Se debe demostrar que W tiene un subespacio complementario T-invariante. De acuerdo con el corolario del Teorema 3, será suficiente demostrar que si f es un polinomio y β es un vector de V tal que $f(T)\beta$ está en W, entonces existe un vector α en W con $f(T)\beta = f(T)\alpha$. Así, pues, supóngase que β está en V y que f es un polinomio tal que $f(T)\beta$ está en W. Si $f(T)\beta = 0$, se hace $\alpha = 0$ y entonces α es un vector en W con $f(T)\beta = f(T)\alpha$. Si $f(T)\beta \neq 0$, el polinomio f no es divisible por el polinomio minimal p del operador T. Como p es primo, eso quiere decir que f y p son primos relativos y existen polinomios g y h tales que $fg + ph = 1$. Como $p(T) = 0$ se tiene entonces que $f(T)g(T) = I$. De esto se sigue que el vector β debe estar en el subespacio W; en efecto,

$$\begin{aligned} \beta &= g(T)f(T)\beta \\ &= g(T)(f(T)\beta) \end{aligned}$$

cuando $f(T)\beta$ está en W y W es invariante por T. Se hace $\alpha = \beta$. $\blacksquare$

Teorema 11. *Sea T un operador lineal sobre un espacio vectorial de dimensión finita V. Una condición necesaria y suficiente para que T sea semisimple*

es que el polinomio minimal p de T sea de la forma $p = p_1 \cdots p_k$, donde $p_1, \ldots, p_k$ son polinomios irreducibles distintos sobre el cuerpo escalar F.

Demostración. Supóngase que T es semisimple. Se demostrará que ningún polinomio irreducible está repetido en la factorización prima del polinomio minimal p. Supóngase lo contrario. Entonces existe algún polinomio mónico g, no escalar, tal que g^2 divide a p. Sea W el espacio nulo del operador $g(T)$. Entonces W es invariante por T. Ahora bien, $p = g^2h$ para algún polinomio h. Como g no es un polinomio escalar, el operador $g(T)h(T)$ no es el operador cero y hay un vector β en V tal que $g(T)h(T)\beta \neq 0$; es decir, $(gh)\beta \neq 0$. Ahora, $(gh)\beta$ está en el subespacio W, ya que $g(gh\beta) = g^2h\beta = p\beta = 0$. Pero no existe vector alguno α en W tal que $gh\beta = gh\alpha$; en efecto, si α está en W

$$(gh)\alpha = (hg)\alpha = h(g\alpha) = h(0) = 0.$$

Así W no puede tener un subespacio complementario T-invariante, contradiciendo la hipótesis de que T es semisimple.

Ahora supóngase que la factorización prima de p es $p = p_1 \cdots p_k$, donde $p_1, \ldots, p_k$ son polinomios mónicos (no escalares) irreducibles distintos. Sea W un subespacio de V invariante por T. Se demostrará que V tiene un subespacio complementario T-invariante. Sea $V = W_1 \oplus \cdots \oplus W_k$ la descomposición prima de T, es decir, sea W_j el espacio nulo de $p_j(T)$. Sea T_j el operador lineal inducido en W_j por T, de modo que el polinomio minimal de T_j sea el primo p_j. Ahora, $W \cap W_j$ es un subespacio de W_j que es invariante por T_j (o por T). Por el último lema existe un subespacio V_j de W_j tal que $W_j = (W \cap W_j) \oplus V_j$, y V_j es invariante por T_j (y, por tanto, por T). Entonces tenemos

$$\begin{aligned} V &= W_1 \oplus \cdots \oplus W_k \\ &= (W \cap W_1) \oplus V_1 \oplus \cdots \oplus (W \cap W_k) \oplus V_k \\ &= (W \cap W_1) + \cdots + (W \cap W_k) \oplus V_1 \oplus \cdots \oplus V_k. \end{aligned}$$

Por el primer lema anterior, $W = (W \cap W_1) \oplus \cdots \oplus (W \cap W_k)$. De modo que si $W' = V_1 \oplus \cdots \oplus V_k$, entonces $V = W \oplus W'$ y W' es invariante por T. ∎

Corolario. *Si T es un operador lineal sobre un espacio vectorial de dimensión finita sobre un cuerpo algebraicamente cerrado, T es semisimple si, y solo si, T es diagonalizable.*

Demostración. Si el cuerpo de los escalares es algebraicamente cerrado, los primos mónicos sobre F son los polinomios $x - c$. En este caso T es semisimple si, y solo si, el polinomio minimal de T es $p = (x - c_1) \cdots (x - c_k)$, donde los $c_1, \ldots, c_k$ son elementos distintos de F. Esto es precisamente el criterio para que T sea diagonalizable, que se estableció en el Capítulo 6. ∎

Debemos señalar que T es semisimple si, y solo si existe un polinomio f, que es producto de factores primos distintos, tal que $f(T) = 0$. Esto es solo en apariencia diferente de la condición de que el polinomio minimal sea un producto de factores primos distintos.

Sea ahora expresar un operador lineal como suma de un operador lineal semisimple y de un operador nilpotente que conmutan. Para ello se restringirá el cuerpo de los escalares a un subcuerpo de los números complejos. El lector advertido verá que lo importante es que el cuerpo F sea un cuerpo de característica cero, esto es, que para todo entero positivo n, la suma $1 + \cdots + 1$ (n veces) en F no debe ser 0. Para un polinomio f sobre F, se representa por $f^{(k)}$ la k-ésima derivada formal de f. En otras palabras, $f^{(k)} = D^k f$, donde D es el operador derivación sobre el espacio de los polinomios. Si g es otro polinomio, $f(g)$ representa el resultado de sustituir g en f, es decir, el polinomio que se obtiene por la aplicación de f al elemento g en el álgebra lineal $F[x]$.

Lema (fórmula de Taylor). *Sea F un cuerpo de característica cero y sean g y h polinomios sobre F. Si f es cualquier polinomio sobre F con* grd $f \leq n$, *entonces*

$$f(g) = f(h) + f^{(1)}(h)(g - h) + \frac{f^{(2)}(h)}{2!}(g - h)^2 + \cdots + \frac{f^{(n)}(h)}{n!}(g - h)^n.$$

Demostración. Lo que estamos demostrando es una generalización de la fórmula de Taylor. El lector probablemente estará familiarizado con el caso especial en que $h = c$, un polinomio escalar, y $g = x$. En tal caso, la fórmula dice que

$$f = f(x) = f(c) + f^{(1)}(c)(x - c) + \frac{f^{(2)}(c)}{2!}(x - c)^2 + \cdots + \frac{f^{(n)}(c)}{n!}(x - c)^n.$$

La demostración de la fórmula general es nada más que una aplicación del teorema del binomio

$$(a + b)^k = a^k + ka^{k-1}b + \frac{k(k-1)}{2!}a^{k-2}b^2 + \cdots + b^k.$$

El lector deberá ver que, como la sustitución y derivación son procesos lineales, solo se necesita demostrar la fórmula cuando $f = x^k$. La fórmula para $f = \sum_{k=0}^{n} c_k x^k$ se deduce por combinación lineal. En el caso $f = x^k$, con $k \leq n$, la fórmula dice que

$$g^k = h^k + kh^{k-1}(g - h) + \frac{k(k-1)}{2!}h^{k-2}(g - h)^2 + \cdots + (g - h)^k$$

que no es más que el desarrollo binómico de

$$g^k = [h + (g - h)]^k. \quad \blacksquare$$

Lema. *Sea F un subcuerpo de los números complejos, sea f un polinomio sobre F y sea f' la derivada de f. Las siguientes afirmaciones son equivalentes:*

(a) *f es el producto de polinomios irreducibles distintos sobre F.*
(b) *f y f' son primos relativos.*

(c) *Como polinomio de coeficientes complejos, f no tiene raíces repetidas.*

Demostración. Demostremos primero que (a) y (b) son afirmaciones equivalentes acerca de f. Supóngase que en la factorización prima de f sobre el cuerpo F, algún polinomio primo (no escalar) p esté repetido. Entonces $f = p^2h$ para algún h en $F[x]$. Entonces

$$f' = p^2h' + 2pp'h$$

y p es también un divisor de f'. Luego f y f' no son primos relativos. Concluimos que (b) implica (a).

Ahora supóngase que $f = p_1 \cdots p_k$, donde $p_1, \ldots, p_k$ son polinomios irreducibles no escalares distintos sobre F. Sea $f_j = f/p_j$. Entonces

$$f' = p_1'f_1 + p_2'f_2 + \cdots + p_k'f_k.$$

Sea p un polinomio primo que divide a f y a f'. Entonces $p = p_i$ para cierto i. Ahora bien, p_i divide a f_j para $j \neq i$, y como p_i también divide a

$$f' = \sum_{j=1}^{k} p_j'f_j$$

vemos que p_i debe dividir a $p_i'f_i$. Por tanto, p_i divide a f_i o a p_i'. Pero p_i no divide a f_i, ya que $p_1, \ldots, p_k$ son distintos. Así p_i divide a p_i'. Esto no es posible, ya que p_i' tiene grado menor en una unidad que el grado de p_i. Concluimos que ningún primo divide a f y a f', o que $(f, f') = 1$.

Para ver que la afirmación (c) es equivalente a (a) y (b), se necesita solamente observar lo siguiente. Supóngase que f y g son polinomios sobre F, un subcuerpo de los números complejos. Se puede considerar también a f y g como polinomios con coeficientes complejos. La afirmación de que f y g son primos relativos como polinomios sobre F, es equivalente a la afirmación de que f y g son primos relativos como polinomios sobre el cuerpo de los números complejos. Se deja la demostración como ejercicio. Se usa este hecho con $g = f'$. Obsérvese que (c) no es más que (a) cuando f es considerado como polinomio sobre el cuerpo de los números complejos. Así (b) y (c) son equivalentes por el mismo razonamiento que se dio anteriormente. ▌

Podemos ahora demostrar un teorema que hace más evidente la relación entre los operadores semisimples y los operadores diagonalizables.

Teorema 12. *Sea F un subcuerpo del cuerpo de los números complejos, sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea T un operador lineal sobre V. Sea $\mathcal{B}$ una base ordenada de V y sea A la matriz de T en la base ordenada $\mathcal{B}$. Entonces T es semisimple si, y solo si, la matriz A es semejante, sobre el cuerpo de los números complejos, a una matriz diagonal.*

Demostración. Sea p el polinomio minimal de T. Según el Teorema 11, T es semisimple si, y solo si, $p = p_1 \cdots p_k$, donde los $p_1, \ldots, p_k$ son polinomios irreducibles distintos sobre F. Por el último lema se ve que T es semisimple si, y solo si, p no tiene raíces complejas repetidas.

Ahora p es también el polinomio minimal de la matriz A. Se sabe que A es semejante, sobre el cuerpo de los números complejos, a una matriz diagonal si, y solo si, su polinomio minimal no tiene raíces complejas repetidas. Esto demuestra el teorema. ▮

Teorema 13. *Sea F un subcuerpo del cuerpo de los números complejos, sea V un espacio vectorial de dimensión finita sobre F y sea T un operador lineal sobre V. Existen un operador semisimple S en V y un operador nilpotente N en V tal que*

(i) $T = S + N$;
(ii) $SN = NS$.

Además, el S semisimple y el N nilpotente que satisfacen (i) *y* (ii), *son únicos y cada uno es un polinomio en T.*

Demostración. Sea $p_1^{r_1} \cdots p^{r_k}$ la factorización prima del polinomio minimal de T y sea $f = p_1 \cdots p_k$. Sea r el mayor de los enteros positivos $r_1, \ldots, r_k$. Entonces el polinomio f es un producto de factores primos distintos, f^r es divisible por el polinomio minimal de T. y así

$$f(T)^r = 0.$$

Se construirá una sucesión de polinomios $g_0, g_1, g_2, \ldots$ tal que

$$f\left(x - \sum_{j=0}^{n} g_j f^j\right)$$

es divisible por f^{n+1}, $n = 0, 1, 2, \ldots$ Se toma $g_0 = 0$ y entonces $f(x - g_0 f^0) = f(x) = f$ es divisible por f. Supóngase que se han elegido los $g_0, \ldots, g_{n-1}$. Sea

$$h = x - \sum_{j=0}^{n-1} g_j f^j$$

de modo que, por suposición, $f(h)$ es divisible por f^n. Se quiere elegir g_n de modo que

$$f(h - g_n f^n)$$

sea divisible por f^{n+1}. Aplicando la fórmula general de Taylor se obtiene

$$f(h - g_n f^n) = f(h) - g_n f^n f'(h) + f^{n+1} b$$

donde b es cierto polinomio. Por hipótesis, $f(h) = qf^n$. Así se ve que para obtener que $f(h - g_n f^n)$ sea divisible por f^{n+1} se necesita solo elegir g_n de tal modo que $(q - g_n f')$ sea divisible por f. Esto puede hacerse, ya que f no tiene factores primos repetidos, con lo que f y f' son primos relativos. Si a y e son polinomios tales que $af + ef' = 1$ y si se hace $g_n = eq$, entonces $q - g_n f'$ es divisible por f.

Se tiene ahora una sucesión $g_0, g_1, \ldots$ tal que f^{n+1} divide a $f\left(x - \sum_{j=0}^{n} g_j f^j\right)$. Tomemos $n = r - 1$, y entonces, como $f(T)^r = 0$,

$$f\left(T - \sum_{j=0}^{r-1} g_j(T)f(T)^j\right) = 0.$$

Sea

$$N = \sum_{j=1}^{r-1} g_j(T)f(T)^j = \sum_{j=0}^{r-1} g_j(T)f(T)^j.$$

Como $\sum_{j=1}^{n} g_j f^j$ es divisible por f, se ve que $N^r = 0$ y N es nilpotente. Sea $S = T - N$, entonces $f(S) = f(T - N) = 0$. Como f tiene factores primos distintos, S es semisimple.

Ahora se tiene $T = S + N$, donde S es semisimple, N nilpotente y cada uno es un polinomio en T. Para demostrar la unicidad, se pasará del cuerpo de los escalares F al cuerpo de los números complejos. Sea $\mathcal{B}$ una base ordenada del espacio V. Se tiene entonces que

$$[T]_{\mathcal{B}} = [S]_{\mathcal{B}} + [N]_{\mathcal{B}}$$

mientras que $[S]_{\mathcal{B}}$ es diagonalizable sobre los números complejos y $[N]_{\mathcal{B}}$ es nilpotente. Estas matrices diagonalizables y nilpotentes que conmutan están unívocamente determinadas, como se vio en el Capítulo 6. ▌

Ejercicios

1. Si N es un operador nilpotente sobre V, demostrar que para cualquier polinomio f, la parte semisimple de $f(N)$ es un múltiplo escalar del operador identidad (F, subcuerpo de C).

2. Sean F un subcuerpo de los números complejos, V un espacio vectorial de dimensión finita sobre F y T un operador lineal semisimple sobre V. Si f es cualquier polinomio sobre F, demostrar que $f(T)$ es semisimple.

3. Sea T un operador lineal sobre un espacio de dimensión finita sobre un subcuerpo de C. Demostrar que T es semisimple si, y solo si, es cierto lo siguiente: si f es un polinomio y $f(T)$ es nilpotente, entonces $f(T) = 0$.

8. Espacios con producto interno

8.1. Productos internos

A lo largo de este capítulo se considerarán solo espacios vectoriales reales o complejos, esto es, espacios vectoriales sobre el cuerpo de los números reales o el cuerpo de los números complejos. El principal objetivo ahora es el estudio de los espacios vectoriales en los que tiene sentido hablar de «longitud» de un vector y de «ángulo» entre dos vectores. Se hará esto mediante el estudio de cierto tipo de función de valor escalar sobre parejas de vectores, conocido como producto interno. Un ejemplo de producto interno es el producto escalar de vectores en R^3. El producto escalar de los vectores

$$\alpha = (x_1, x_2, x_3) \quad \text{y} \quad \beta = (y_1, y_2, y_3)$$

en R^3 es el número real

$$(\alpha|\beta) = x_1y_1 + x_2y_2 + x_3y_3.$$

Geométricamente, este producto escalar es el producto de la longitud de α, la longitud de β y el coseno del ángulo entre α y β. Es, por tanto, posible definir los conceptos geométricos de «longitud» y «ángulo» en R^3 por la definición algebraica de un producto escalar.

Un producto interno sobre un espacio vectorial es una función con propiedades similares a las del producto escalar en R^3, y en términos de tal producto interno se puede también definir «longitud» y «ángulo». El comentario respecto a la noción general de ángulo se restringirá al concepto de perpendicularidad (u ortogonalidad) de vectores. En esta primera sección se dirá qué es un producto interno, se considerarán unos ejemplos particulares y se establecerán

algunas de las propiedades básicas de los productos internos. Entonces se volverá a la tarea de discutir longitud y ortogonalidad.

Definición. *Sean F el cuerpo de los números reales, o de los complejos, y V un espacio vectorial sobre F. Un* **producto interno** *sobre V es una función que asigna a cada par ordenado de vectores* α, β *de V un escalar* $(\alpha|\beta)$ *de F de tal modo que para cualesquiera* α, β, γ *de V y todos los escalares c*

(a) $(\alpha + \beta|\gamma) = (\alpha|\gamma) + (\beta|\gamma)$;
(b) $(c\alpha|\beta) = c(\alpha|\beta)$;
(c) $(\beta|\alpha) = \overline{(\alpha|\beta)}$, *donde la barra indica conjugación compleja*;
(d) $(\alpha|\alpha) > 0$ *si* $\alpha \neq 0$.

Debe observarse que las condiciones (a), (b) y (c) implican que

(e) $$(\alpha|c\beta + \gamma) = \bar{c}(\alpha|\beta) + (\alpha|\gamma).$$

Otra observación debe hacerse. Si F es el cuerpo R de los números reales, los complejos conjugados que figuran en (c) y (e) están demás; sin embargo, en el caso de los complejos son necesarias para la consistencia de las condiciones. Sin estos complejos conjugados se tendría la contradicción:

$$(\alpha|\alpha) > 0 \quad \text{y} \quad (i\alpha|i\alpha) = -1(\alpha|\alpha) > 0.$$

En los ejemplos que siguen y a lo largo del capítulo, F es el cuerpo de los números reales o de los números complejos.

Ejemplo 1. En F^n existe un producto interno que se llama **producto interno canónico.** Está definido sobre $\alpha = (x_1, \ldots, x_n)$ y $\beta = (y_1, \ldots, y_n)$ por

(8-1) $$(\alpha|\beta) = \sum_j x_j\bar{y}_j.$$

Cuando $F = R$, esto también puede escribirse

$$(\alpha|\beta) = \sum_j x_j y_j.$$

En el caso real, el producto interno canónico es el llamado a menudo producto escalar, y se representa por $\alpha \cdot \beta$.

Ejemplo 2. Para $\alpha = (x_1, x_2)$ y $\beta = (y_1, y_2)$ en R^2, sea

$$(\alpha|\beta) = x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2.$$

Como $(\alpha|\alpha) = (x_1 - x_2)^2 + 3x_2^2$, se tiene que $(\alpha|\alpha) > 0$ si $\alpha \neq 0$. Las condiciones (a), (b) y (c) de la definición se pueden verificar fácilmente.

Ejemplo 3. Sean $V = F^{n \times n}$, el espacio de todas las matrices $n \times n$ sobre F. Entonces V es isomorfo a F^{n^2} de un modo natural; se sigue, pues, del Ejemplo 1, que la igualdad

$$(A|B) = \sum_{j,k} A_{jk}\bar{B}_{jk}$$

define un producto interno sobre V. Además, si se introduce la **matriz transpuesta conjugada** B^*, donde $B^*_{kj} = \bar{B}_{jk}$, se puede expresar este producto interno sobre $F^{n \times n}$ en términos de la función traza:

$$(A|B) = \operatorname{tr}(AB^*) = \operatorname{tr}(B^*A).$$

En efecto,

$$\begin{aligned} \operatorname{tr}(AB^*) &= \sum_j (AB^*)_{jj} \\ &= \sum_j \sum_k A_{jk}B^*_{kj} \\ &= \sum_j \sum_k A_{jk}\bar{B}_{jk}. \end{aligned}$$

Ejemplo 4. Sea $F^{n \times 1}$ el espacio de las matrices (columnas) $n \times 1$ sobre F, y sea Q una matriz inversible $n \times n$ sobre F. Para X, Y en $F^{n \times 1}$ se hace

$$(X|Y) = Y^*Q^*QX.$$

Se identifica la matriz 1×1 del segundo miembro con su solo elemento. Cuando Q es la matriz unidad, este producto interno es esencialmente el mismo del Ejemplo 1; se llama el **producto interno canónico** sobre $F^{n \times 1}$. El lector deberá observar que la terminología de «producto interno canónico» se usa en dos contextos especiales. Para un espacio vectorial general de dimensión finita sobre F no hay un producto interno obvio que se pueda llamar canónico.

Ejemplo 5. Sea V el espacio vectorial de las funciones continuas de valor complejo en el intervalo unitario, $0 \leq t \leq 1$. Sea

$$(f|g) = \int_0^1 f(t)\overline{g(t)}\,dt.$$

El lector está probablemente más familiarizado con el espacio de las funciones de valor real en el intervalo unitario, y para este espacio la conjugación compleja puede omitirse.

Ejemplo 6 Este es en realidad un conjunto de ejemplos. Por el siguiente método se pueden construir nuevos productos internos a partir de uno dado. Sean V y W espacios vectoriales sobre F y supóngase que $(\ |\ )$ es un producto interno sobre W. Si T es una transformación lineal no singular de V en W, entonces la igualdad

$$p_T(\alpha, \beta) = (T\alpha|T\beta)$$

define un producto interno p_T sobre V. El producto interno del Ejemplo 4 es un caso especial de esta situación. También lo son los siguientes:

(a) Sea V un espacio vectorial de dimensión finita y sea

$$\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}$$

una base ordenada de V. Sean $\epsilon_1, \ldots, \epsilon_n$ los vectores de la base canónica en F^n y sea T la transformación lineal de V en F^n tal que $T\alpha_j = \epsilon_j$, $j = 1, \ldots, n$. En otras palabras, sea T el isomorfismo «natural» de V sobre el producto interno canónico sobre F^n, entonces

$$p_T(\sum_j x_j\alpha_j, \sum_k y_k\alpha_k) = \sum_{j=1}^{n} x_j\bar{y}_j.$$

Así, para cada base para V existe un producto interno sobre V con la propiedad de que $(\alpha_j|\alpha_k) = \delta_{jk}$; en efecto, es fácil ver que existe exactamente uno solo de tales productos internos. Más adelante se verá que cada producto interno sobre V está determinado por alguna base $\mathcal{B}$ en la forma anterior.

(b) Volvamos nuevamente al Ejemplo 5 y tomemos $V = W$, el espacio de las funciones continuas en el intervalo unitario. Sea T el operador lineal «multiplicación por t», esto es, $(Tf)(t) = tf(t)$, $0 \leq t \leq 1$. Es fácil ver que T es lineal. También T es no singular; en efecto, supóngase que $Tf = 0$. Entonces $tf(t) = 0$ para $0 \leq t \leq 1$; luego $f(t) = 0$ para $t > 0$. Como f es continua, se tiene también que $f(0) = 0$, o $f = 0$. Ahora, usando el producto interno del Ejemplo 5, se construye un producto interno nuevo sobre V haciendo

$$\begin{aligned} p_T(f, g) &= \int_0^1 (Tf)(t)\overline{(Tg)(t)}\, dt \\ &= \int_0^1 f(t)\overline{g(t)}t^2\, dt. \end{aligned}$$

Se vuelve ahora a las observaciones generales para los productos internos. Supóngase que V es un espacio vectorial complejo con un producto interno. Entonces, para todo α, β en V

$$(\alpha|\beta) = \text{Re}\,(\alpha|\beta) + i\,\text{Im}\,(\alpha|\beta)$$

donde Re $(\alpha|\beta)$ e Im $(\alpha|\beta)$ son las partes real e imaginaria del número complejo $(\alpha|\beta)$. Si z es un número complejo, entonces Im (z) = Re $(-iz)$. Se sigue que

$$\text{Im}\,(\alpha|\beta) = \text{Re}\,[-i(\alpha|\beta)] = \text{Re}\,(\alpha|i\beta).$$

Así, el producto interno está completamente determinado por su «parte real» de acuerdo con

(8-2) $$(\alpha|\beta) = \text{Re}\,(\alpha|\beta) + i\,\text{Re}\,(\alpha|i\beta).$$

Ocasionalmente es muy útil saber que un producto interno sobre un espacio vectorial, real o complejo, está determinado por otra función, la llamada forma cuadrática determinada por el producto interno. Para definirla, primero se representa la raíz cuadrada positiva de $(\alpha|\alpha)$ por $||\alpha||$; $||\alpha||$ es la llamada **norma** de α respecto al producto interno. Considerando el producto interno canónico en R^1, C^1, R^2 y R^3, el lector podrá convencerse de que es apropiado pensar en la norma de α como la «longitud» o «magnitud» de α. La **forma cuadrática** determinada por el producto interno es la función que asigna a cada vector α el escalar $||\alpha||^2$. Se sigue de las propiedades del producto interno que

$$||\alpha \pm \beta||^2 = ||\alpha||^2 \pm 2\,\text{Re}\,(\alpha|\beta) + ||\beta||^2$$

para todos los vectores α y β. Así, en el caso real

(8-3) $$(\alpha|\beta) = \frac{1}{4}||\alpha + \beta||^2 - \frac{1}{4}||\alpha - \beta||^2.$$

En el caso complejo se usa (8-2) para obtener la expresión más complicada

$$(8\text{-}4)\qquad (\alpha|\beta) = \frac{1}{4}\|\alpha+\beta\|^2 - \frac{1}{4}\|\alpha-\beta\|^2 + \frac{i}{4}\|\alpha+i\beta\|^2 - \frac{i}{4}\|\alpha-i\beta\|^2.$$

Las igualdades (8-3) y (8-4) son llamadas las **identidades de polarización.** Obsérvese que (8-4) puede también ser escrita de la siguiente manera:

$$(\alpha|\beta) = \frac{1}{4}\sum_{n=1}^{4} i^n \|\alpha + i^n\beta\|^2.$$

Las propiedades obtenidas anteriormente son válidas para cualquier producto interno sobre un espacio real o complejo V, independientemente de su dimensión. Se vuelve al caso en que V es de dimensión finita. Como es de pensarlo, un producto interno en un espacio de dimensión finita puede ser descrito siempre en términos de una base ordenada por medio de una matriz.

Supóngase que V es de dimensión finita, de modo que

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$$

es una base ordenada de V y que se tiene dado un producto interno particular sobre V; se verá que el producto interno está completamente determinado por los valores

$$(8\text{-}5)\qquad G_{jk} = (\alpha_k|\alpha_j)$$

que tienen los pares de vectores en $\mathcal{B}$. Si $\alpha = \sum_k x_k\alpha_k$ y $\beta = \sum_j y_j\alpha_j$, entonces

$$\begin{aligned}(\alpha|\beta) &= (\sum_k x_n\alpha_k|\beta)\\ &= \sum_k x_k(\alpha_k|\beta)\\ &= \sum_k x_k \sum_j \bar{y}_j(\alpha_k|\alpha_j)\\ &= \sum_{j,k} \bar{y}_j G_{jk} x_k\\ &= Y^*GX\end{aligned}$$

donde X, Y son las matrices de coordenadas de α y β en la base ordenada $\mathcal{B}$, y G es la matriz con elementos $G_{jk} = (\alpha_k|\alpha_j)$. Se llama a G la **matriz del producto interno en la base ordenada** $\mathcal{B}$. Se sigue de (8-5) que G es hermítica, es decir, que $G = G^*$; sin embargo, G es una clase de matriz hermítica más bien especial. En efecto, G debe satisfacer, además, la condición

$$(8\text{-}6)\qquad X^*GX > 0, \qquad X \neq 0.$$

En particular, G debe ser inversible. Ya que en caso contrario existe una $X \neq 0$ tal que $GX = 0$, y para toda tal X, (8-6) es imposible. En forma más explícita. (8-6) dice que para los escalares $x_1, \ldots, x_n$, no todos nulos,

$$(8\text{-}7)\qquad \sum_{j,k} \bar{x}_j G_{jk} x_k > 0.$$

Con esto se observa inmediatamente que cada elemento de la diagonal de G debe ser positivo; sin embargo, esta condición para los elementos de la diagonal no es del todo suficiente para asegurar la validez de (8-6). Condiciones suficientes para la validez de (8-6) se darán más adelante.

El proceso anterior es reversible; esto es, si G es cualquier matriz $n \times n$ sobre F que satisface (8-6) y la condición de que $G = G^*$, entonces G es la matriz en la base ordenada $\mathcal{B}$ de un producto interno sobre V. Este producto interno está dado por

$$(\alpha|\beta) = Y^*GX$$

donde X e Y son las matrices de coordenadas de α y β en la base ordenada $\mathcal{B}$.

Ejercicios

1. Sea V un espacio vectorial y $(\ |\ )$ un producto interno sobre V.
(a) Demostrar que $(0|\beta) = 0$ para todo β de V.
(b) Demostrar que si $(\alpha|\beta) = 0$ para todo β de V, entonces $\alpha = 0$.

2. Sea V un espacio vectorial sobre F. Demostrar que la suma de dos productos internos sobre V es un producto interno sobre V. ¿Es la diferencia de dos productos internos un producto interno? Mostrar que un múltiplo positivo de un producto interno es un producto interno.

3. Describir explícitamente todos los productos internos sobre R^1 y sobre C^1.

4. Comprobar que el producto interno canónico sobre F^n es un producto interno.

5. Sea $(\ |\ )$ el producto interno canónico sobre R^2.
(a) Sean $\alpha = (1, 2)$, $\beta = (-1, 1)$. Si γ es un vector tal que $(\alpha|\gamma) = -1$ y $(\beta|\gamma) = 3$, hallar γ.
(b) Demostrar que para cada α en R^2 se tiene que $\alpha = (\alpha|\epsilon_1)\epsilon_1 + (\alpha|\epsilon_2)\epsilon_2$.

6. Sea $(\ |\ )$ el producto interno canónico sobre R^2 y sea T el operador lineal $T(x_1, x_2) = (-x_2, x_1)$. Ahora, T es la «rotación en 90°» y tiene la propiedad de que $(\alpha|T\alpha) = 0$ para todo α en R^2. Hallar todos los productos internos $[\ |\ ]$ en R^2 tales que $[\alpha|T\alpha] = 0$ para cada α.

7. Sea $(\ |\ )$ el producto interno canónico sobre C^2. Demostrar que existe un operador lineal no nulo en C^2 tal que $(\alpha|T\alpha) = 0$ para cada α en C^2. Generalizarlo.

8. Sea A una matriz 2×2 con elementos reales. Para X, Y en $F^{2\times 1}$, sea

$$f_A(X, Y) = Y^tAX.$$

Demostrar que f_A es un producto interno sobre $R^{2\times 1}$ si, y solo si, $A = A^t$, $A_{11} > 0$, $A_{22} > 0$ y $\det A > 0$.

9. Sea V un espacio vectorial real o complejo con un producto interno. Demostrar que la forma cuadrática determinada por el producto interno cumple la **ley del paralelogramo**

$$||\alpha + \beta||^2 + ||\alpha - \beta||^2 = 2||\alpha||^2 + 2||\beta||^2.$$

10. Sea $(\ |\ )$ el producto interno sobre R^2 definido en el Ejemplo 2 y sea $\mathcal{B}$ la base ordenada canónica de R^2. Hallar la matriz de este producto interno con respecto a $\mathcal{B}$.

11. Mostrar que la fórmula

$$\left(\sum_j a_j x^j \middle| \sum_k b_k x^k\right) = \sum_{j,k} \frac{a_j b_k}{j + k + 1}$$

define un producto interno sobre el espacio $R[x]$ de los polinomios sobre el cuerpo R. Sea W el subespacio de los polinomios de grado menor o igual que n. Restringir el producto interno anterior a W y hallar la matriz de este producto interno sobre W respecto a la base ordenada $\{1, x, \ldots, x^n\}$. (*Sugerencia*: Para mostrar que la fórmula define un producto interno, obsérvese que

$$(f|g) = \int_0^1 f(t)g(t)\, dt$$

y operar con la integral.)

12. Sea V un espacio vectorial de dimensión finita y sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base de V. Sea $(\ |\ )$ un producto interno sobre V. Si $c_1, \ldots, c_n$ son n escalares arbitrarios, demostrar que existe exactamente un vector α en V tal que $(\alpha|\alpha_j) = c_j$, $j = 1, \ldots, n$.

13. Sea V un espacio vectorial complejo. Una función J de V en V se llama si $J(\alpha + \beta) = J(\alpha) + J(\beta)$, $J(c\alpha) = cJ(\alpha)$ y $J(J(\alpha)) = \alpha$ para todo c y todo α, β en V. Si J es una conjugación, demostrar que

(a) El conjunto W de todos los α en V tales que $J\alpha = \alpha$ es un espacio vectorial sobre R con respecto a las operaciones definidas en V.

(b) Para cada α en V existen vectores únicos β, γ en W tales que $\alpha = \beta + i\gamma$.

14. Sean V un espacio vectorial complejo y W un subconjunto de V con las siguientes propiedades:

(a) W es un espacio vectorial real con respecto a las operaciones definidas en V.

(b) Para cada α en V existen vectores únicos β, γ en W tales que $\alpha = \beta + i\gamma$. Demostrar que la ecuación $J\alpha = \beta - i\gamma$ define una conjugación en V tal que $J\alpha = \alpha$ si, y solo si, α pertenece a W y mostrar también que J es la única conjugación en V con esta propiedad.

15. Hallar todas las conjugaciones en C^1 y C^2.

16. Sea W un subespacio real de dimensión finita de un espacio vectorial complejo V. Demostrar que W satisface la condición (b) del Ejercicio 14 si, y solo si, cada base de W es también una base de V.

17. Sean V un espacio vectorial complejo, J una conjugación sobre V, W el conjunto de los α de V tales que $J\alpha = \alpha$ y f un producto interno sobre W. Demostrar que:

(a) Existe un producto interno único g en V tal que $g(\alpha, \beta) = f(\alpha, \beta)$ para todo α, β en W.

(b) $g(J\alpha, J\beta) = g(\beta, \alpha)$ para todo α, β en V.

¿Qué dice la parte (a) respecto a la relación entre los productos internos canónicos sobre R^1 y C^1 o sobre R^n y C^n?

8.2. *Espacios producto interno*

Ahora que se tiene una idea de lo que es un producto interno, se atenderá a lo que se puede decir al respecto de la combinación de un espacio vectorial y un producto interno particular sobre el mismo. Específicamente se estable-

cerán las propiedades básicas de los conceptos de «longitud» y «ortogonalidad» que son impuestos al espacio por el producto interno.

Definición. *Un* **espacio producto interno** *es un espacio real o complejo junto con un producto interno definido sobre ese espacio.*

Un espacio producto interno real de dimensión finita se llama a menudo **espacio euclidiano.** Un espacio con producto interno complejo se llama a veces **espacio unitario.**

Teorema 1. *Si V es un espacio producto interno, entonces para vectores α, β cualesquiera de V y cualquier escalar c*

(i) $\|c\alpha\| = |c|\ \|\alpha\|$;
(ii) $\|\alpha\| > 0$ *para* $\alpha \neq 0$;
(iii) $|(\alpha|\beta)| \leq \|\alpha\|\ \|\beta\|$;
(iv) $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$.

Demostración. Las afirmaciones (i) y (ii) se desprenden casi inmediatamente de las varias definiciones vistas. La desigualdad en (iii) es claramente válida para $\alpha = 0$. Si $\alpha \neq 0$, hágase

$$\gamma = \beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha.$$

Entonces $(\gamma|\alpha) = 0$ y

$$\begin{aligned} 0 \leq \|\gamma\|^2 &= \left(\beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha \,\middle|\, \beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha\right) \\ &= (\beta|\beta) - \frac{(\beta|\alpha)(\alpha|\beta)}{\|\alpha\|^2} \\ &= \|\beta\|^2 - \frac{|(\alpha|\beta)|^2}{\|\alpha\|^2}. \end{aligned}$$

Luego $|(\alpha|\beta)|^2 \leq \|\alpha\|^2\|\beta\|^2$. Ahora usando (c) se tiene que

$$\begin{aligned} \|\alpha + \beta\|^2 &= \|\alpha\|^2 + (\alpha|\beta) + (\beta|\alpha) + \|\beta\|^2 \\ &= \|\alpha\|^2 + 2\,\mathrm{Re}\,(\alpha|\beta) + \|\beta\|^2 \\ &\leq \|\alpha\|^2 + 2\,\|\alpha\|\,\|\beta\| + \|\beta\|^2 \\ &= (\|\alpha\| + \|\beta\|)^2. \end{aligned}$$

con lo que $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$. ∎

La desigualdad en (iii) se llama **desigualdad de Cauchy-Schwarz** y tiene una amplia variedad de aplicaciones. La demostración muestra que si (por ejemplo) α es cero, entonces $|(\alpha|\beta)| \leq \|\alpha\|\ \|\beta\|$ a menos que

$$\beta = \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha.$$

Así, la igualdad se tiene en (iii) si, y solo si, α y β son linealmente dependientes.

Ejemplo 7. Si se aplica la desigualdad de Cauchy-Schwarz al producto interno dado en los Ejemplos 1, 2, 3 y 5 se obtiene lo siguiente

(a) $$|\Sigma\, x_k\bar{y}_k| \leq (\Sigma\, |x_k|^2)^{1/2}(\Sigma\, |y_k|^2)^{1/2}$$

(b) $$|x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2| \leq ((x_1 - x_2)^2 + 3x_2^2)^{1/2}((y_1 - y_2)^2 + 3y_2^2)^{1/2}$$

(c) $$|\text{tr}\,(AB^*)| \leq (\text{tr}\,(AA^*))^{1/2}(\text{tr}\,(BB^*))^{1/2}$$

(d) $$\left|\int_0^1 f(x)\overline{g(x)}\, dx\right| \leq \left(\int_0^1 |f(x)|^2\, dx\right)^{1/2}\left(\int_0^1 |g(x)|^2\, dx\right)^{1/2}.$$

Definiciones. *Sean α y β vectores de un espacio producto interno V. Entonces α es* **ortogonal** *a β si $(\alpha|\beta) = 0$; como esto implica que β es ortogonal a α, a menudo solo se dirá que α y β son ortogonales. Si S es un conjunto de vectores de V, se dice que S es un* **conjunto ortogonal** *siempre que todos los pares de vectores distintos de S sean ortogonales. Un* **conjunto ortonormal** *es un conjunto ortogonal S con la propiedad además de que $\|\alpha\| = 1$ para todo α de S.*

El vector cero es ortogonal a todo vector en V y es el único vector con esa propiedad. Es apropiado pensar un conjunto ortonormal como conjunto de vectores mutuamente perpendiculares que tienen longitud 1.

Ejemplo 8. La base canónica en R^n o C^n es un conjunto ortonormal con respecto al producto interno canónico.

Ejemplo 9. El vector (x, y) de R^2 es ortogonal a $(-y, x)$ con respecto al producto interno canónico; en efecto,

$$((x, y)|(-y, x)) = -xy + yx = 0.$$

Sin embargo, si R^2 está dotado del producto interno del Ejemplo 2, entonces (x, y) y $(-y, x)$ son ortogonales si, y solo si,

$$y = \tfrac{1}{2}\,(-3 \pm \sqrt{13})x.$$

Ejemplo 10. Sea $V = C^{n\times n}$ el espacio de las matrices complejas $n \times n$, y sea $E^{p,q}$ la matriz cuyo único elemento no nulo es un 1 en la fila p y la columna q. Entonces el conjunto de todas las matrices E^{pq} es ortonormal con respecto al producto interno dado en el Ejemplo 3. En efecto,

$$(E^{pq}|E^{rs}) = \text{tr}\,(E^{pq}E^{sr}) = \delta_{qs}\,\text{tr}\,(E^{pr}) = \delta_{qs}\delta_{pr}.$$

Ejemplo 11. Sea V el espacio de las funciones continuas de valor complejo (o valor real) en el intervalo $0 \leq x \leq 1$ con el producto interno

$$(f|g) = \int_0^1 f(x)\overline{g(x)}\, dx.$$

Supóngase que $f_n(x) = \sqrt{2}\,\cos 2\pi nx$ y que $g_n(x) = \sqrt{2}\,\text{sen}\, 2\pi nx$. Entonces

$\{1, f_1, g_1, f_2, g_2, \ldots\}$ es un conjunto infinito ortonormal. En el caso complejo se pueden formar también las combinaciones lineales

$$\frac{1}{\sqrt{2}}(f_n + ig_n), \qquad n = 1, 2, \ldots.$$

De este modo se tiene un nuevo conjunto ortonormal S que consta de todas las funciones de la forma

$$h_n(x) = e^{2\pi inx}, \qquad n = \pm 1, \pm 2, \ldots.$$

El conjunto S' que se obtiene de S adjuntando la función constante 1 es también ortonormal. Se supone aquí que el lector está familiarizado con el cálculo de las integrales en cuestión.

Los conjuntos ortogonales dados en los ejemplos anteriores son todos linealmente independientes. Se ve ahora que necesariamente éste es el caso.

Teorema 2. *Un conjunto ortogonal de vectores no nulos es linealmente independiente.*

Demostración. Sea S un conjunto ortogonal finito o infinito de vectores no nulos en un espacio con producto interno dado. Supóngase que $\alpha_1, \alpha_2, \ldots, \alpha_m$ son vectores distintos en S y que

$$\beta = c_1\alpha_1 + c_2\alpha_2 + \cdots + c_m\alpha_m.$$

Entonces

$$\begin{aligned}(\beta|\alpha_k) &= (\sum_j c_j\alpha_j|\alpha_k)\\ &= \sum_j c_j(\alpha_j|\alpha_k)\\ &= c_k(\alpha_k|\alpha_k).\end{aligned}$$

Como $(\alpha_k|\alpha_k) \neq 0$, se sigue que

$$c_k = \frac{(\beta|\alpha_k)}{||\alpha_k||^2}, \qquad 1 \leq k \leq m.$$

Así, cuando $\beta = 0$, cada $c_k = 0$, de modo que S es un conjunto independiente. ∎

Corolario. *Si un vector β es combinación lineal de una sucesión ortogonal de vectores no nulos $\alpha_1, \ldots, \alpha_m$, entonces β es igual a la combinación lineal particular*

(8-8)
$$\beta = \sum_{k=1}^{m} \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\alpha_k.$$

Este corolario se desprende de la demostración del teorema. Hay otro corolario que mencionar que es también obvio. Si $\{\alpha_1, \ldots, \alpha_m\}$ es un conjunto ortogonal de vectores no nulos en un espacio con producto interno de dimensión finita V, entonces $m \leq \dim V$. Esto dice que el número de direcciones mutuamente perpendiculares en V no pueden exceder la dimensión algebraica-

mente definida de V. El máximo número de direcciones mutuamente perpendiculares en V es lo que intuitivamente se considera como dimensión geométrica de V, y se acaba de ver que ésta no es mayor que la dimensión algebraica. El que estas dos dimensiones sean iguales es un corolario particular del siguiente resultado.

Teorema 3. *Sea V un espacio con producto interno y sean $\beta_1, \ldots, \beta_n$ vectores independientes cualesquiera de V. Entonces se pueden construir vectores ortogonales $\alpha_1, \ldots, \alpha_n$ en V tales que para cada $k = 1, 2, \ldots, n$ el conjunto*

$$\{\alpha_1, \ldots, \alpha_k\}$$

sea una base del subespacio generado por $\beta_1, \ldots, \beta_k$.

Demostración. Los vectores $\alpha_1, \ldots, \alpha_n$ se obtendrán por medio de una construcción conocida como **proceso de ortogonalización de Gram-Schmidt.** Primero, sea $\alpha_1 = \beta_1$. Los otros vectores están entonces dados inductivamente de la siguiente forma: supóngase que $\alpha_1, \ldots, \alpha_m$ $(1 \leq m \leq n)$ hayan sido elegidos de modo que para cada k

$$\{\alpha_1, \ldots, \alpha_k\}, \qquad 1 \leq k \leq m$$

es una base ortogonal para el subespacio de V que es generado por $\beta_1, \ldots, \beta_k$. Para construir el siguiente vector α_{m+1}, sea

$$\alpha_{m+1} = \beta_{m+1} - \sum_{k=1}^{m} \frac{(\beta_{m+1}|\alpha_k)}{||\alpha_k||^2}\alpha_k. \tag{8-9}$$

Entonces $\alpha_{m+1} \neq 0$. En efecto, ya que en caso contrario α_{m+1} es una combinación lineal de los $\alpha_1, \ldots, \alpha_m$ y luego una combinación lineal de los $\alpha_1, \ldots, \alpha_m$. Además, si $1 \leq j \leq m$, entonces

$$\begin{aligned}(\alpha_{m+1}|\alpha_j) &= (\beta_{m+1}|\alpha_j) - \sum_{k=1}^{m} \frac{(\beta_{m+1}|\alpha_k)}{||\alpha_k||^2}(\alpha_k|\alpha_j)\\ &= (\beta_{m+1}|\alpha_j) - (\beta_{m+1}|\alpha_j)\\ &= 0.\end{aligned}$$

Por tanto, $\{\alpha_1, \ldots, \alpha_{m+1}\}$ es un conjunto ortogonal que consta de $m + 1$ vectores no nulos en el subespacio generado por $\beta_1, \ldots, \beta_{m+1}$. Por el Teorema 2, él es una base para este subespacio. Así los vectores $\alpha_1, \ldots, \alpha_n$ pueden construirse uno tras otro de acuerdo con (8-9). En particular, cuando $n = 4$, se tiene

$$\begin{aligned}\alpha_1 &= \beta_1\\ \alpha_2 &= \beta_2 - \frac{(\beta_2|\alpha_1)}{||\alpha_1||^2}\alpha_1\\ \alpha_3 &= \beta_3 - \frac{(\beta_3|\alpha_1)}{||\alpha_1||^2}\alpha_1 - \frac{(\beta_3|\alpha_2)}{||\alpha_2||^2}\alpha_2\\ \alpha_4 &= \beta_4 - \frac{(\beta_4|\alpha_1)}{||\alpha_1||^2}\alpha_1 - \frac{(\beta_4|\alpha_2)}{||\alpha_2||^2}\alpha_2 - \frac{(\beta_4|\alpha_3)}{||\alpha_3||^2}\alpha_3.\end{aligned} \tag{8-10}$$ ∎

Corolario. *Todo espacio con producto interno de dimensión finita tiene una base ortonormal.*

Demostración. Sean V un espacio con producto interior de dimensión finita y $\{\beta_1, \ldots, \beta_n\}$ una base para V. Aplicando el proceso de Gram-Schmidt, se construye una base ortogonal $\{\alpha_1, \ldots, \alpha_n\}$. Entonces, para obtener una base ortonormal, se remplaza simplemente cada vector α_k por $\alpha_k/\|\alpha_k\|$. ∎

Una de las ventajas principales que tienen las bases ortonormales sobre las bases arbitrarias es que los cálculos en que entran coordenadas son más simples. Para indicar en términos generales por qué es así, supóngase que V es un espacio con producto interior de dimensión finita. Entonces, como en la última sección, se puede usar (8-5) para asociar la matriz G con cada base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ de V. Usando esta matriz

$$G_{jk} = (\alpha_k|\alpha_j),$$

se puede calcular el producto interno en términos de las coordenadas. Si $\mathcal{B}$ es una base ortonormal, entonces G es la matriz unidad, y para cualesquiera escalares x_j e y_k

$$(\sum_j x_j\alpha_j|\sum_k y_k\alpha_k) = \sum_j x_j\bar{y}_j.$$

Con lo que, en términos de una base ortonormal, el producto interno en V se asemeja al producto interno canónico en F^n.

Aunque es poco útil para los cálculos, es interesante observar que el proceso de Gram-Schmidt puede ser también usado como criterio de independencia lineal. En efecto, supóngase que $\beta_1, \ldots, \beta_n$ son vectores linealmente dependientes en un espacio con producto interno V. Para excluir el caso trivial, supóngase que $\beta_1 \neq 0$. Sea m el mayor entero para el que los $\beta_1, \ldots, \beta_m$ son independientes. Entonces $1 \leq m \leq n$. Sean $\alpha_1, \ldots, \alpha_m$ los vectores obtenidos por la aplicación del proceso de ortogonalización a los $\beta_1, \ldots, \beta_m$. Entonces el vector α_{m+1} dado por (8-9) es necesariamente 0. En efecto, α_{m+1} está en el subespacio generado por $\alpha_1, \ldots, \alpha_m$ y es ortogonal a cada uno de estos vectores; luego es 0 por (8-8). Recíprocamente, si $\alpha_1, \ldots, \alpha_m$ son distintos de 0 y $\alpha_{m+1} = 0$, entonces $\beta_1, \ldots, \beta_{m+1}$ son linealmente dependientes.

Ejemplo 12. Se consideran los vectores

$$\begin{aligned}\beta_1 &= (3, 0, 4)\\ \beta_2 &= (-1, 0, 7)\\ \beta_3 &= (2, 9, 11)\end{aligned}$$

en R^3 con el producto interno canónico. Aplicando el proceso de Gram-Schmidt a los $\beta_1, \beta_2, \beta_3$, se obtienen los siguientes vectores

$$\begin{aligned}\alpha_1 &= (3, 0, 4)\\ \alpha_2 &= (-1, 0, 7) - \frac{((-1, 0, 7)|(3, 0, 4))}{25}(3, 0, 4)\\ &= (-1, 0, 7) - (3, 0, 4)\\ &= (-4, 0, 3)\end{aligned}$$

$$\alpha_3 = (2, 9, 11) - \frac{((2, 9, 11)|(3, 0, 4))}{25}(3, 0, 4)$$
$$- \frac{((2, 9, 11)|(-4, 0, 3))}{25}(-4, 0, 3)$$
$$= (2, 9, 11) - 2(3, 0, 4) - -4, 0, 3)$$
$$= (0, 9, 0).$$

Estos vectores son evidentemente no nulos y mutuamente perpendiculares. Luego $\{\alpha_1, \alpha_2, \alpha_3\}$ es una base ortogonal para R^3. Para expresar un vector arbitrario (x_1, x_2, x_3) en R^3 como combinación lineal de los $\alpha_1, \alpha_2, \alpha_3$ *no* es necesario resolver ninguna ecuación lineal. Para ello es suficiente usar (8-8). Así,

$$(x_1, x_2, x_3) = \frac{3x_1 + 4x_3}{25}\alpha_1 + \frac{-4x_1 + 3x_3}{25}\alpha_2 + \frac{x_2}{9}\alpha_3$$

como se verifica rápidamente. En particular,

$$(1, 2, 3) = \tfrac{3}{5}(3, 0, 4) + \tfrac{1}{5}(-4, 0, 3) + \tfrac{2}{9}(0, 9, 0).$$

Para ver esto desde otro punto de vista, lo que se tiene es lo siguiente: la base $\{f_1, f_2, f_3\}$ de $(R^3)^*$, que es dual de la base $\{\alpha_1, \alpha_2, \alpha_3\}$, está definida explícitamente por:

$$f_1(x_1, x_2, x_3) = \frac{3x_1 + 4x_3}{25}$$

$$f_2(x_1, x_2, x_3) = \frac{-4x_1 + 3x_3}{25}$$

$$f_3(x_1, x_2, x_3) = \frac{x_2}{9}$$

y estas ecuaciones pueden ser escritas en forma más general como

$$f_j(x_1, x_2, x_3) = \frac{((x_1, x_2, x_3)|\alpha_j)}{||\alpha_j||^2}.$$

Finalmente, obsérvese que de $\alpha_1, \alpha_2, \alpha_3$ se tiene la base ortonormal

$$\tfrac{1}{5}(3, 0, 4), \quad \tfrac{1}{5}(-4, 0, 3), \quad (0, 1, 0).$$

Ejemplo 13. Sea $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, donde a, b, c y d son números complejos. Se hace $\beta_1 = (a, b)$, $\beta_2 = (c, d)$ y se supone que $\beta_1 \neq 0$. Si se aplica el proceso de ortogonalización a β_1, β_2, usando el producto interno canónico en C^2, se obtienen los siguientes vectores

$$\alpha_1 = (a, b)$$

$$\begin{aligned}\alpha_2 &= (c, d) - \frac{((c, d)|(a, b))}{|a|^2 + |b|^2}(a, b)\\ &= (c, d) - \frac{(c\bar{a} + d\bar{b})}{|a|^2 + |b|^2}(a, b)\\ &= \left(\frac{cb\bar{b} - d\bar{b}a}{|a|^2 + |b|^2}, \frac{d\bar{a}a - c\bar{a}b}{|a|^2 + |b|^2}\right)\\ &= \frac{\det A}{|a|^2 + |b|^2}(-\bar{b}, \bar{a}).\end{aligned}$$

Ahora la teoría general dice que $\alpha_2 \neq 0$ si, y solo si, β_1, β_2 son linealmente independientes. Por otro lado, la fórmula para α_2 muestra que este es el caso si, y solo si, $\det A \neq 0$.

En esencia, el proceso de Gram-Schmidt consiste en la aplicación repetida de una operación geométrica básica llamada proyección ortogonal, y se le entiende mejor desde este punto de vista. El método de la proyección ortogonal también aparece en forma natural en la solución de un importante problema de aproximación.

Supóngase que W es un subespacio de un espacio con producto interno V, y sea β un vector arbitrario en V. El problema consiste en hallar una mejor aproximación posible a β por vectores de W. Esto quiere decir que se desea encontrar un vector α para el que $||\beta - \alpha||$ sea lo más pequeño posible, sujeto a la restricción de que α debe pertenecer a W. Precisemos lo dicho.

Una **mejor aproximación** a β por vectores de W es un vector α de W tal que

$$||\beta - \alpha|| \leq ||\beta - \gamma||$$

para todo vector γ en W.

Observando este problema en R^2, o en R^3, se ve intuitivamente que una mejor aproximación a β por vectores de W debe ser un vector α de W tal que $\beta - \alpha$ es perpendicular (ortogonal) a W y que tal α debe ser único. Estas ideas intuitivas son correctas para subespacios de dimensión finita y para algunos, pero no todos, los subespacios de dimensión infinita. Dado que la situación precisa es demasiado complicada de tratar aquí, se demostrará solo el siguiente resultado.

Teorema 4. *Sea W un subespacio de un espacio producto interno V y sea β un vector de V.*

(i) *El vector α en W es una mejor aproximación a β, por vectores de W si, y solo si, $\beta - \alpha$ es ortogonal a todo vector de W.*

(ii) *Si existe una mejor aproximación a β por vectores de W es única.*

(iii) *Si W es de dimensión finita y $\{\alpha_1, \ldots, \alpha_n\}$ es cualquier base ortonormal de W, entonces el vector*

$$\alpha = \sum_k \frac{(\beta|\alpha_k)}{||\alpha_k||^2}\alpha_k$$

es la (única) mejor aproximación a β por vectores de W.

Demostración. Primeramente, obsérvese que si γ es cualquier vector de V, entonces $\beta - \gamma = (\beta - \alpha) + (\alpha - \gamma)$, y

$$||\beta - \gamma||^2 = ||\beta - \alpha||^2 + 2 \operatorname{Re} (\beta - \alpha|\alpha - \gamma) + ||\alpha - \gamma||^2.$$

Supóngase ahora que $\beta - \alpha$ es ortogonal a todo vector de W, que γ está en W y que $\gamma \neq \alpha$. Entonces, como $\alpha - \gamma$ está en W, se sigue que

$$\begin{aligned} ||\beta - \gamma||^2 &= ||\beta - \alpha||^2 + ||\alpha - \gamma||^2 \\ &> ||\beta - \alpha||^2. \end{aligned}$$

Recíprocamente, supóngase que $||\beta - \gamma|| \geq ||\beta - \alpha||$ para todo γ en E. Entonces de la primera ecuación anterior se sigue que

$$2 \operatorname{Re} (\beta - \alpha|\alpha - \gamma) + ||\alpha - \gamma||^2 \geq 0$$

para todo γ de W. Como todo vector de W puede ser expresado en la forma $\alpha - \gamma$ con γ en W, se ve que

$$2 \operatorname{Re} (\beta - \alpha|\tau) + ||\tau||^2 \geq 0$$

para todo τ en W. En particular, si γ está en W y $\gamma \neq \alpha$, se puede tener

$$\tau = -\frac{(\beta - \alpha|\alpha - \gamma)}{||\alpha - \gamma||^2} (\alpha - \gamma).$$

Entonces, la desigualdad se reduce a la afirmación

$$-2 \frac{|(\beta - \alpha|\alpha - \gamma)|^2}{||\alpha - \gamma||^2} + \frac{|(\beta - \alpha|\alpha - \gamma)|^2}{||\alpha - \gamma||^2} \geq 0.$$

Lo cual se cumple si, y solo si, $(\beta - \alpha|\alpha - \gamma) = 0$. Por tanto, $\beta - \alpha$ es ortogonal a todo vector en W. Esto completa la demostración de la equivalencia de las dos condiciones para α dadas en (i). La condición de ortogonalidad es evidentemente satisfecha por, a lo más, un vector de W, lo que demuestra (ii).

Ahora supóngase que W es un subespacio de dimensión finita de V. Entonces se sabe, por un corolario del Teorema 3, que W tiene una base ortogonal. Sea $\{\alpha_1, \ldots, \alpha_n\}$ cualquier base ortogonal para W, y se define α por (8-11). Entonces, por el cálculo hecho en la demostración del Teorema 3, $\beta - \alpha$ es ortogonal a cada uno de los vectores α_k ($\beta - \alpha$ es el vector obtenido en la última etapa cuando se aplica el proceso de ortogonalización a $\alpha_1, \ldots, \alpha_n, \beta$). Así $\beta - \alpha$ es ortogonal a toda combinación lineal de los $\alpha_1, \ldots, \alpha_n$, es decir, a todo vector en W. Si γ está en W y $\gamma \neq \alpha$, se sigue que $||\beta - \gamma|| > ||\beta - \alpha||$. Por tanto, α es la mejor aproximación a β que está en W.

Definición. *Sea V un espacio producto interno y S cualquier conjunto de vectores en V. El* **complemento ortogonal** *de S es el conjunto $S^\perp$ de los vectores de V ortogonales a todo vector de S.*

El complemento ortogonal de V es el subespacio cero y, recíprocamente, $\{0\}^\perp = V$. Si S es cualquier subconjunto de V, su complemento ortogonal $S^\perp$ (S perp) es siempre un subespacio de V. En efecto, como $S^\perp$ no es vacío, por

contener al 0, y toda vez que α y β están en $S^{\perp}$ y c es escalar cualquiera, se tiene

$$\begin{aligned}(c\alpha + \beta|\gamma) &= c(\alpha|\gamma) + (\beta|\gamma)\\ &= c0 + 0\\ &= 0\end{aligned}$$

para cada γ en S, así $c\alpha + \beta$ también está en $S^{\perp}$. En el Teorema 4, la propiedad característica del vector α es que es el único vector de W tal que $\beta - \alpha$ pertenece a $W^{\perp}$

Definición. *Siempre que exista el vector α en el Teorema 4 se le llama* **proyección ortogonal de β sobre W**. *Si todo vector de V tiene proyección ortogonal sobre W, la aplicación que asigna a cada vector de V su proyección ortogonal sobre W, se llama* **proyección ortogonal de V sobre W**.

Por el Teorema 4, la proyección ortogonal de un espacio producto interno sobre un subespacio de dimensión finita siempre existe. Pero el Teorema 4 también implica el siguiente resultado.

Corolario. *Sean V un espacio producto interno, W un subespacio de dimensión finita y E la proyección ortogonal de V sobre W. Entonces la aplicación*

$$\beta \rightarrow \beta - E\beta$$

es la proyección ortogonal de V sobre $W^{\perp}$.

Demostración. Sea β un vector arbitrario de V. Entonces $\beta - E\beta$ está en W y para cualquier γ en $W^{\perp}$, $\beta - \gamma = E\beta + (\beta - E\beta - \gamma)$. Como $E\beta$ está en W y $\beta - E\beta - \gamma$ está en $W^{\perp}$, se sigue que

$$\begin{aligned}||\beta - \gamma||^2 &= ||E\beta||^2 + ||\beta - E\beta - \gamma||^2\\ &\geq ||\beta - (\beta - E\beta)||^2\end{aligned}$$

que hace más estricta la desigualdad cuando $\gamma \neq \beta - E\beta$. Por tanto, $\beta - E\beta$ es la mejor aproximación a β para vectores en $W^{\perp}$.

Ejemplo 14. Se da a R^3 el producto interno canónico. Entonces la proyección ortogonal de $(-10, 2, 8)$ sobre el subespacio W generado por $(3, 12, -1)$ es el vector

$$\begin{aligned}\alpha &= \frac{((-10, 2, 8)|(3, 12, -1))}{9 + 144 + 1}(3, 12, -1)\\ &= \frac{-14}{154}(3, 12, -1).\end{aligned}$$

La proyección ortogonal de R^3 sobre W es la transformación lineal E definida por

$$(x_1, x_2, x_3) \rightarrow \left(\frac{3x_1 + 12x_2 - x_3}{154}\right)(3, 12, -1).$$

La imagen de E es evidentemente 1; luego su nulidad es 2. Por otro lado.

$$E(x_1, x_2, x_3) = (0, 0, 0)$$

si, y solo si, $3x_1 + 12x_2 = x_3 = 0$. Este es el caso si, y solo si, (x_1, x_2, x_3) está en $W^\perp$. Por tanto, $W^\perp$ es el espacio nulo de E, y $\dim (W^\perp) = 2$. Calculando

$$(x_1, x_2, x_3) - \left(\frac{3x_1 + 12x_2 - x_3}{154}\right)(3, 12, -1)$$

se ve que la proyección ortogonal de R^3 sobre $W^\perp$ es la transformación lineal $I - E$ que aplica el vector (x_1, x_2, x_3) sobre el vector

$$\frac{1}{154}(145x_1 - 36x_2 + 3x_3, -36x_1 + 10x_2 + 12x_3, 3x_1 + 12x_2 + 153x_3).$$

Las observaciones hechas en el Ejemplo 14 se generalizan de la siguiente forma.

Teorema 5. *Sea W un subespacio de dimensión finita de un espacio producto interno V y sea E la proyección ortogonal de V sobre W. Entonces E es una transformación lineal idempotente de V sobre W, $W^\perp$ es el espacio nulo de E y*

$$V = W \oplus W^\perp.$$

Demostración. Sea β un vector arbitrario en V. Entonces $E\beta$ es la mejor aproximación a β que está en W. En particular, $E\beta = \beta$ cuando β está en W. Por tanto, $E(E\beta) = E\beta$ para todo β en V; esto es, E es idempotente: $E^2 = E$. Para demostrar que E es una transformación lineal, sean α y β vectores cualesquiera de V y c un escalar arbitrario. Entonces, por el Teorema 4, $\alpha - E\alpha$ y $\beta - E\beta$ son ambos ortogonales a todo vector de W. Luego el vector

$$c(\alpha - E\alpha) + (\beta - E\beta) = (c\alpha + \beta) - (cE\alpha + E\beta)$$

también pertenece a $W^\perp$. Como $cE\alpha + E\beta$ es un vector de W, se sigue del Teorema 4 que

$$E(c\alpha + \beta) = cE\alpha + E\beta.$$

Por cierto que se puede también demostrar la linealidad de E usando (8-11). Nuevamente, sea β un vector cualquiera de V. Entonces E es el vector único de W tal que $\beta - E\beta$ está en $W^\perp$. Así, $E\beta = 0$ cuando β está en $W^\perp$. Recíprocamente, β está en $W^\perp$ cuando $E\beta = 0$. Así $W^\perp$ es el espacio nulo de E. La ecuación

$$\beta = E\beta + \beta - E\beta$$

muestra que $V = W + W^\perp$; más aún, $W \cap W^1 = \{0\}$. En efecto, si α es un vector de $W \cap W^\perp$, entonces $(\alpha|\alpha) = 0$. Por tanto, $\alpha = 0$ y V es la suma directa de W y $W^\perp$. ∎

Corolario. *Bajo las condiciones del teorema, $I - E$ es la proyección ortogonal de V en $W^\perp$. El es una transformación lineal idempotente de V en $W^\perp$ con espacio nulo W.*

Demostración. Acabamos de ver que la aplicación $\beta \rightarrow \beta - E\gamma$ es la proyección ortogonal de V sobre $W^\perp$. Como E es una transformación lineal, esta proyección en $W^\perp$ es la proyección lineal $I - E$. Por sus propiedades geométricas se ve que $I - E$ es una transformación idempotente de V sobre W. Esto también se desprende del cálculo

$$\begin{aligned}(I - E)(I - E) &= I - E - E + E^2 \\ &= I - E.\end{aligned}$$

Además, $(I - E)\beta = 0$ si, y solo si, $\beta = E\beta$, y este es el caso si, y solo si, β está en W. Por tanto, W es el subespacio nulo de $I - E$. ∎

El proceso de Gram-Schmidt puede ser ahora descrito geométricamente del siguiente modo. Dado un espacio producto interno V y los vectores $\beta_1, \ldots, \beta_n$ en V, sea P_k $(k > 1)$ la proyección ortogonal de V en el complemento ortogonal del subespacio generado por $\beta_1, \ldots, \beta_{k-1}$ y sea $P_1 = I$. Entonces, los vectores que se obtienen por aplicación del proceso de ortogonalización a $\beta_1, \ldots, \beta_n$ están definidos por las ecuaciones

$$\alpha_k = P_k\beta_k, \qquad 1 \leq k \leq n. \tag{8-12}$$

El Teorema 5 implica otro resultado conocido como **desigualdad de Bessel.**

Corolario. *Sea* $\{\alpha_1, \ldots, \alpha_n\}$ *un conjunto ortogonal de vectores no nulos en un espacio producto interno* V. *Si* β *es cualquier vector de* V, *entonces*

$$\sum_k \frac{|(\beta|\alpha_k)|^2}{\|\alpha_k\|^2} \leq \|\beta\|^2$$

y la desigualdad vale si, y solo si,

$$\beta = \sum_k \frac{(\beta|\alpha_k)}{\|\alpha_k\|^2}\alpha_k.$$

Demostración. *Sea* $\gamma = \sum\limits_k [(\beta|\alpha_k)/\|\alpha_k\|^2]\,\alpha_k$. Entonces $\beta = \gamma + \delta$, donde $(\gamma|\delta) = 0$. Luego

$$\|\beta\|^2 = \|\gamma\|^2 + \|\delta\|^2.$$

Es ahora suficiente demostrar que

$$\|\gamma\|^2 = \sum_k \frac{|(\beta|\alpha_k)|^2}{\|\alpha_k\|^2}.$$

Este es el cálculo rutinario en el que se usa el hecho de que $(\alpha_j|\alpha_k) = 0$ para $j \neq k$. ∎

En el caso especial en que $\{\alpha_1, \ldots, \alpha_n\}$ es un conjunto ortonormal, la desigualdad de Bessel dice que

$$\sum_k |(\beta|\alpha_k)|^2 \leq \|\beta\|^2.$$

El corolario dice también, en este caso, que β está en el subespacio generado por $\{\alpha_1, \ldots, \alpha_n\}$ si, y solo si,

$$\beta = \sum_k (\beta|\alpha_k)\, \alpha_k$$

o si, y solo si, la desigualdad de Bessel es efectivamente una igualdad. Por cierto que en el caso en que V sea de dimensión finita y $\{\alpha_1, \ldots, \alpha_n\}$ es una base ortogonal de V, la fórmula anterior rige para todo vector β en V. En otras palabras; si $\{\alpha_1, \ldots, \alpha_n\}$ es una base ortonormal de V, la k-ésima coordenada de β en la base ordenada $\{\alpha_1, \ldots, \alpha_n\}$ es $(\beta|\alpha_k)$.

Ejemplo 15. Se aplicará el último corolario al conjunto ortogonal descrito en el Ejemplo 11. Hallamos que

(a) $$\sum_{k=-n}^{n} \left| \int_0^1 f(t) e^{-2\pi ikt}\, dt \right|^2 \leq \int_0^1 |f(t)|^2\, dt$$

(b) $$\int_0^1 \left| \sum_{k=-n}^{n} c_k e^{2\pi ikt} \right|^2 dt = \sum_{k=-n}^{n} |c_k|^2$$

(c) $$\int_0^1 (\sqrt{2} \cos 2\pi t + \sqrt{2} \operatorname{sen} 4\pi t)^2\, dt = 1 + 1 = 2.$$

Ejercicios

1. Considerar R^4, con el producto interno canónico. Sea W el subespacio de R^4 que consta de todos los vectores ortogonales a $\alpha = (1, 0, -1, 1)$ y $\beta = (2, 3, -1, 2)$. Hallar una base de W.

2. Aplicar el proceso de Gram-Schmidt a los vectores $\beta_1 = (1, 0, 1)$, $\beta_2 = (1, 0, -1)$, $\beta_3 = (0, 3, 4)$ para obtener una base ortogonal para R^3 con el producto interno canónico.

3. Considerar C^3 con el producto interno canónico. Hallar una base ortonormal para el subespacio generado por $\beta_1 = (1, 0, i)$ y $\beta_2 = (2, 1, 1 + i)$.

4. Sea V un espacio producto interno. La *distancia* entre los vectores α y β en V está definida por

$$d(\alpha, \beta) = ||\alpha - \beta||.$$

Demostrar que

(a) $d(\alpha, \beta) \geq 0$;
(b) $d(\alpha, \beta) = 0$ si, y solo si, $\alpha = \beta$;
(c) $d(\alpha, \beta) = d(\beta, \alpha)$;
(d) $d(\alpha, \beta) \leq d(\alpha, \gamma) + d(\gamma, \beta)$.

5. Sea V un espacio producto interno y sean α, β vectores de V. Demostrar que $\alpha = \beta$ si, y solo si, $(\alpha|\gamma) = (\beta|\gamma)$ para todo γ de V.

6. Sea W el subespacio de R^2 generado por el vector (3, 4). Usando el producto interno canónico, sea E la proyección ortogonal de R^2 sobre W. Hallar
(a) una fórmula para $E(x_1, x_2)$;
(b) la matriz de E en la base ordenada canónica;
(c) $W^\perp$;
(d) una base ortonormal en que E está representada por la matriz

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

7. Sea V el espacio producto interno que consta de R^2 y el producto interno cuya forma cuadrática está definida por

$$||(x_1, x_2||^2 = (x_1 - x_2)^2 + 3x_2^2.$$

Sea E la proyección ortogonal de V sobre el subespacio W generado por el vector (3, 4). Responder ahora las cuatro preguntas del Ejercicio 6.

8. Hallar un producto interno sobre R^2 tal que $(\epsilon_1, \epsilon_2) = 2$.

9. Sea V el subespacio de polinomios $R[x]$ con grado a lo más 3. Dótese V con el producto interno

$$(f|g) = \int_0^1 f(t)g(t)\, dt.$$

(a) Hallar el complemento ortogonal del subespacio de polinomios escalares;
(b) Aplicar el proceso de Gram-Schmidt a la base $\{1, x, x^2, x^3\}$.

10. Sea V el espacio vectorial de todas las matrices $n \times n$ sobre C, con el producto interno $(A|B) = \operatorname{tr}(AB^*)$. Hallar el complemento ortogonal del subespacio de matrices diagonales.

11. Sea V un espacio producto interno de dimensión finita y sea $\{\alpha_1, \ldots, \alpha_n\}$ una base ortonormal de V. Demostrar que para vectores α, β cualesquiera de V

$$(\alpha|\beta) = \sum_{k=1}^{n} (\alpha|\alpha_k)\overline{(\beta|\alpha_k)}.$$

12. Sea W un subespacio de dimensión finita de un espacio producto interno V y sea E la proyección ortogonal de V sobre W. Demostrar que $(E\alpha|\beta) = (\alpha|E\beta)$ para todo α, β en V.

13. Sea S un subconjunto de un espacio producto interno V. Demostrar que $(S^\perp)^\perp$ contiene al subespacio generado por S. Cuando V es de dimensión finita, mostrar que $(S^\perp)^\perp$ es el subespacio generado por S.

14. Sea V un espacio producto interno de dimensión finita y sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ortonormal de V. Sea T un operador lineal sobre V y A la matriz de T en la base ordenada $\mathcal{B}$. Demostrar que

$$A_{ij} = (T\alpha_j|\alpha_i).$$

15. Supóngase que $V = W_1 \oplus W_2$ y que f_1 y f_2 son productos internos en W_1 y W_2, respectivamente. Demostrar que existe un producto interno único f sobre V tal que
(a) $W_2 = W_1^\perp$;
(b) $f(\alpha, \beta) = f_k(\alpha, \beta)$, cuando α, β están en W_k, $k = 1, 2$.

16. Sea V un espacio producto interno y W un subespacio de V de dimensión finita. Existen (en general) muchas proyecciones que tienen a W como imagen. Una de éstas, la pro-

yección ortogonal sobre W, tiene la propiedad de que $\|E\alpha\| \leq \|\alpha\|$ para todo α de V. Demostrar que si E es una proyección con imagen W, tal que $\|E\alpha\| \leq \|\alpha\|$ para todo α de V, entonces E es la proyección ortogonal sobre W.

17. Sea V el espacio producto interno real que consta del espacio de las funciones continuas de valor real en el intervalo $-1 \leq t \leq 1$, con el producto interno.

$$(f|g) = \int_{-1}^{1} f(t)g(t)\,dt.$$

Sea W el subespacio de las funciones impares, es decir, las funciones que satisfacen $f(-t) = -f(t)$. Hallar el complemento ortogonal de W.

8.3. *Funciones lineales y adjuntas*

La primera parte de esta sección se dedicará a los funcionales lineales sobre un espacio producto interno. El resultado principal es que cualquier funcional lineal f sobre un espacio producto interno de dimensión finita es «un producto interno con un vector fijo en el espacio»; es decir, que tal f tiene la forma $f(\alpha) = (\alpha|\beta)$ para algún β fijo de V. Se usará este resultado para demostrar la existencia del «adjunto» de un operador lineal T sobre V, siendo éste un operador lineal T^* tal que $(T\alpha|\beta) = (\alpha|T^*\beta)$ para todo α y β en V. Por el uso de una base ortonormal, esta operación adjunta sobre operadores lineales (que pasa de T a T^*) se identifica con la operación de formar la transpuesta conjugada de una matriz. Exploramos someramente la analogía entre la operación adjunta y la conjugación de números complejos.

Sea V cualquier espacio producto interno y sea β un vector fijo de V. Se define una función f_β de V en el cuerpo escalar por

$$f_\beta(\alpha) = (\alpha|\beta).$$

Esta función f_β es un funcional lineal sobre V, ya que por su propia definición, $(\alpha|\beta)$ es lineal como función de α. Si V es de dimensión finita, todo funcional lineal sobre V se puede expresar de esta forma para cierto β.

Teorema 6. *Sean V un espacio producto interno de dimensión finita y f un funcional lineal sobre V. Entonces, existe un único vector β de V tal que $f(\alpha) = (\alpha|\beta)$ para todo α de V.*

Demostración. Sea $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ una base ortonormal de V. Se hace

(8-13) $$\beta = \sum_{j=1}^{n} \overline{f(\alpha_j)}\alpha_j$$

y sea f_β el funcional lineal definido por

$$f_\beta(\alpha) = (\alpha|\beta).$$

Entonces

$$f_\beta(\alpha_k) = (\alpha_k|\sum_j \overline{f(\alpha_j)}\alpha_j) = f(\alpha_k).$$

Como esto es cierto para todo α_k, se sigue que $f = f_\beta$. Ahora supóngase que γ es un vector de V para el que $(\alpha|\beta) = (\alpha|\gamma)$ para todo α. Entonces $(\beta - \gamma|\beta - \gamma) = 0$ y $\beta = \gamma$. Así existe exactamente un vector β que determina al funcional lineal f en la forma establecida. ▌

La demostración de este teorema puede alterarse ligeramente en el caso de la representación de los funcionales lineales en una base. Si se elige una base ortonormal $\{\alpha_1, \ldots, \alpha_n\}$ para V, el producto interno de $\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n$ y $\beta = y_1\alpha_1 + \cdots + y_n\alpha_n$ será

$$(\alpha|\beta) = x_1\bar{y}_1 + \cdots + x_n\bar{y}_n.$$

Si f es un funcional lineal cualquiera sobre V, entonces f tiene la forma

$$f(\alpha) = c_1x_1 + \cdots + c_nx_n$$

para escalares fijos $c_1, \ldots, c_n$ determinados por la base. Ciertamente, $c_j = f(\alpha_j)$. Si se desea encontrar un vector β en V tal que $(\alpha|\beta) = f(\alpha)$ para todo α, entonces es evidente que las coordenadas y_j de β deben satisfacer $\bar{y}_j = c_j$ o $y_j = \overline{f(\alpha_j)}$. Por consiguiente,

$$\beta = \overline{f(\alpha_1)}\alpha_1 + \cdots + \overline{f(\alpha_n)}\alpha_n$$

es el vector que se desea.

Unos comentarios al respecto. La demostración del Teorema 6 que se ha dado es muy breve, pero no resalta el hecho geométrico esencial de que β está en el complemento ortogonal del espacio nulo de f. Sea W el espacio nulo de f. Entonces $V = W + W^\perp$ y f está completamente determinado por sus valores en $W^\perp$. En efecto, si P es la proyección ortogonal de V sobre $W^\perp$, entonces

$$f(\alpha) = f(P\alpha)$$

para todo α de V. Supóngase que $f \neq 0$. Entonces f es el rango 1 y dim $(W^\perp) = 1$. Si γ es otro vector no nulo cualquiera de $W^\perp$, se sigue que

$$P\alpha = \frac{(\alpha|\gamma)}{||\gamma||^2}\gamma$$

para todo α en V. Así,

$$f(\alpha) = (\alpha|\gamma) \cdot \frac{f(\gamma)}{||\gamma||^2}$$

para todo α, y $\beta = [\overline{f(\gamma)}/||\gamma||^2]\,\gamma$.

Ejemplo 16. He aquí un ejemplo que muestra que el Teorema 6 no es cierto sin la suposición de que V es de dimensión finita. Sea V el espacio vectorial de los polinomios sobre el cuerpo de los números complejos, con el producto interno

$$(f|g) = \int_0^1 f(t)\overline{g(t)}\,dt.$$

Este producto interno también puede ser definido algebraicamente. Si $f = \Sigma\, a_k x^k$ y $g = \Sigma\, b_k x^k$, entonces

$$(f|g) = \sum_{j,k} \frac{1}{j + k + 1}\, a_j \overline{b_k}.$$

Sea z un número complejo fijo y sea L el funcional lineal «evaluación en z»:

$$L(f) = f(z).$$

¿Existe un polinomio g tal que $(f|g) = L(f)$ para todo f? La respuesta es no; en efecto, supóngase que tenemos que

$$f(z) = \int_0^1 f(t)\overline{g(t)}\, dt$$

para todo f. Sea $h = x - z$, de modo que para cualquier f se tenga que $(hf) = 0$. Entonces

$$0 = \int_0^1 h(t)f(t)\overline{g(t)}\, dt$$

para todo f. En particular, esto rige cuando $f = \bar{h}g$, con lo que

$$\int_0^1 |h(t)|^2|g(t)|^2\, dt = 0$$

y así $hg = 0$. Como $h \neq 0$, debe tenerse que $g = 0$. Pero L no es el funcional cero; luego no existe tal g.

Se puede generalizar algo el ejemplo para el caso donde L sea combinación lineal de evaluaciones puntuales. Supóngase que se eligen números complejos fijos $z_1, \ldots, z_n$ y escalares $c_1, \ldots, c_n$, y sea

$$L(f) = c_1 f(z_1) + \cdots + c_n f(z_n).$$

Entonces L es un funcional lineal en V, pero no existe g tal que $L(f) = (f|g)$, al menos que $c_1 = c_2 = \cdots = c_n = 0$. Solo debe repetirse el razonamiento anterior con $h = (x - z_1) \cdots (x - z_n)$.

Se vuelve ahora al concepto del adjunto de un operador lineal.

Teorema 7. *Para cualquier operador lineal T en un espacio producto interno de dimensión finita, existe un único operador lineal T^* sobre V tal que*

$$(8\text{-}14) \qquad (T\alpha|\beta) = (\alpha|T^*\beta)$$

para todo α, β de V.

Demostración. Sea β un vector cualquiera de V. Entonces $\alpha \rightarrow (T\alpha|\beta)$ es un funcional lineal sobre V. Por el Teorema 6 existe un único vector β' en V tal que $(T\alpha|\beta) = (\alpha|\beta')$ para todo α en V. Sea T^* la aplicación $\beta \rightarrow \beta'$:

$$\beta' = T^*\beta.$$

Se tiene (8-14), pero se debe verificar también que T^* es un operador lineal. Sean β, γ en V y sea c un escalar. Entonces para cualquier α,

$$\begin{aligned}(\alpha|T^*(c\beta + \gamma)) &= (T\alpha|c\beta + \gamma)\\ &= (T\alpha|c\beta) + (T\alpha|\gamma)\\ &= \bar{c}(T\alpha|\beta) + (T\alpha|\gamma)\\ &= \bar{c}(\alpha|T^*\beta) + (\alpha|T^*\gamma)\\ &= (\alpha|cT^*\beta) + (\alpha|T^*\gamma)\\ &= (\alpha|cT^*\beta + T^*\gamma).\end{aligned}$$

Así $T^*(c\beta + \gamma) = cT^*\beta + T^*\gamma$, y T^* es lineal.

La unicidad de T^* es inmediata. En efecto, para cualquier β en V, el vector $T^*\beta$ está unívocamente determinado como el vector β' tal que $(T\alpha|\beta) = (\alpha|\beta')$ para todo α. ∎

Teorema 8. *Sea V un espacio producto interno de dimensión finita y sea* $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ *una base (ordenada) ortonormal de V. Sea T un operador lineal sobre V y sea A la matriz de T en la base ordenada* $\mathcal{B}$. *Entonces* $A_{kj} = (T\alpha_j|\alpha_k)$.

Demostración. Como $\mathcal{B}$ es una base ortonormal, se tiene que

$$\alpha = \sum_{k=1}^{n} (\alpha|\alpha_k)\alpha_k.$$

La matriz A está definida por

$$T\alpha_j = \sum_{k=1}^{n} A_{kj}\alpha_k$$

y como

$$T\alpha_j = \sum_{k=1}^{n} (T\alpha_j|\alpha_k)\alpha_k$$

se tiene $A_{kj} = (T\alpha_j|\alpha_k)$. ∎

Corolario. *Sea V un espacio producto interno de dimensión finita y sea T un operador lineal sobre V. En cualquier base ortogonal de V la matriz de* T^* *es la conjugada de la transpuesta de la matriz de T.*

Demostración. Sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ortonormal de V, sea $A = [T]_{\mathcal{B}}$ y $B = [T^*]_{\mathcal{B}}$. De acuerdo con el Teorema 8,

$$A_{kj} = (T\alpha_j|\alpha_k)$$

$$B_{kj} = (T^*\alpha_j|\alpha_k).$$

Por definición de T^* se tiene que

$$\begin{aligned}B_{kj} &= (T^*\alpha_j|\alpha_k)\\ &= \overline{(\alpha_k|T^*\alpha_j)}\\ &= \overline{(T\alpha_k|\alpha_j)}\\ &= \overline{A_{jk}}.\end{aligned}$$ ∎

Ejemplo 17. Sea V un espacio producto interno de dimensión finita y E la proyección ortogonal de V sobre un subespacio W. Entonces para los vectores α y β cualesquiera en V.

$$\begin{aligned}(E\alpha|\beta) &= (E\alpha|E\beta + (1 - E)\beta)\\ &= (E\alpha|E\beta)\\ &= (E\alpha + (1 - E)\alpha|E\beta)\\ &= (\alpha|E\beta).\end{aligned}$$

De la unicidad del operador E^* se sigue que $E^* = E$. Ahora considérese la proyección E descrita en el Ejemplo 14. Entonces

$$A = \frac{1}{154}\begin{bmatrix} 9 & 36 & -3\\ 36 & 144 & -12\\ -3 & -12 & 1\end{bmatrix}$$

es la matriz de E en la base ortonormal canónica. Como $E = E^*$, A es también la matriz de E^*, y como $A = A^*$ esto no contradice el corolario anterior. Por otro lado, supóngase que

$$\begin{aligned}\alpha_1 &= (154, 0, 0)\\ \alpha_2 &= (145, -36, 3)\\ \alpha_3 &= (-36, 10, 12).\end{aligned}$$

Entonces $\{\alpha_1, \alpha_2, \alpha_3\}$ es una base, y

$$\begin{aligned}E\alpha_1 &= (9, 36, -3)\\ E\alpha_2 &= (0, 0, 0)\\ E\alpha_3 &= (0, 0, 0).\end{aligned}$$

Como $(9, 36, -3) = -(154, 0, 0) - (145, -36, 2)$, la matriz B de E en la base $\{\alpha_1, \alpha_2, \alpha_3\}$ está definida por la expresión:

$$B = \begin{bmatrix} -1 & 0 & 0\\ -1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}.$$

En este caso $B \neq B^*$ y B^* no es la matriz de $E^* = E$ en la base $\{\alpha_1, \alpha_2, \alpha_3\}$. Aplicando el corolario, se concluye que $\{\alpha_1, \alpha_2, \alpha_3\}$ no es una base ortonormal. Esto es bastante obvio naturalmente.

Definición. *Sea T un operador lineal sobre un espacio producto interno V. Entonces se dice que T tiene un* **adjunto** *sobre V si existe un operador lineal T^* sobre V tal que $(T\alpha|\beta) = (\alpha|T^*\beta)$ para todo α y β en V.*

Por el Teorema 7 todo operador lineal en un espacio con producto interno de dimensión finita V tiene un adjunto en V. En el caso de dimensión infinita, esto no es siempre cierto. Pero, en todo caso, existe a lo más un tal operador T^*; cuando existe se le llama el **adjunto de** T.

Dos comentarios respecto al caso de dimensión finita.

1. El adjunto de T depende no solo de T, sino también del producto interno.

2. Como se vio en el Ejemplo 17, en una base ordenada arbitraria $\mathcal{B}$, la relación entre $[T]_{\mathcal{B}}$ y $[T^*]_{\mathcal{B}}$ es más complicada que la dada en el corolario anterior.

Ejemplo 18. Sea $V = C^{n\times 1}$ el espacio de las matrices complejas $n \times 1$ con el producto interno $(X|Y) = Y^*X$. Si A es una matriz $n \times n$ con elementos complejos, el adjunto del operador lineal $X \to AX$ es el operador $X \to A^*X$. En efecto,

$$(AX|Y) = Y^*AX = (A^*Y)^*X = (X|A^*Y).$$

El lector debe convencerse por sí mismo de que esto es realmente un caso especial del último corolario.

Ejemplo 19. Este es semejante al Ejemplo 18. Sea $V = C^{n\times n}$ con el producto interno $(A|B) = \text{tr}\,(B^*A)$. Sea M una matriz $n \times n$ dada sobre C. El adjunto de la multiplicación a la izquierda por M es multiplicación a la izquierda por M^*. Naturalmente, «multiplicación a la izquierda por M» es el operador lineal L_M definido por $L_M(A) = MA$.

$$\begin{aligned}(L_M(A)|B) &= \text{tr}\,(B^*(MA))\\ &= \text{tr}\,(MAB^*)\\ &= \text{tr}\,(AB^*M)\\ &= \text{tr}\,(A(M^*B)^*)\\ &= (A|L_M{}^*(B)).\end{aligned}$$

Así $(L_M)^* = L_{M^*}$. En el cálculo anterior se usó dos veces la propiedad característica de la función traza: $\text{tr}\,(AB) = \text{tr}\,(BA)$.

Ejemplo 20. Sea V el espacio de los polinomios sobre el cuerpo de los números complejos con el producto interno

$$(f|g) = \int_0^1 f(t)\overline{g(t)}\,dt.$$

Si f es un polinomio $f = \Sigma\, a_k x^k$, se hace $\bar{f} = \Sigma\, \bar{a}_k x^k$. Esto es, $\bar{f}$ es el polinomio cuya función polinomio asociada es la compleja conjugada de la de f:

$$\bar{f}(t) = \overline{f(t)}, \qquad t \text{ real}$$

Se considera el operador «multiplicación por f», esto es, el operador lineal M_f definido por $(M_f)(g) = fg$. Entonces este operador tiene un adjunto, a saber, multiplicación por $\bar{f}$. En efecto,

$$\begin{aligned}(M_f(g)|h) &= (fg|h)\\ &= \int_0^1 f(t)g(t)\overline{h(t)}\,dt\\ &= \int_0^1 g(t)[\overline{\overline{f(t)}h(t)}]\,dt\\ &= (g|\bar{f}h)\\ &= (g|M_{\bar{f}}(h))\end{aligned}$$

y así $(M_{\bar{f}})^* = M_f$.

Ejemplo 21. En el Ejemplo 20 se vio que algunos operadores en un espacio producto interno de dimensión infinita tiene un adjunto. Como se comentó anteriormente, otros no lo tienen. Sea V el espacio con producto interno del Ejemplo 20 y sea D el operador derivación en $C[x]$. Integrando por partes se tiene que

$$(Df|g) = f(1)g(1) - f(0)g(0) - (f|Dg).$$

Se considera g fijo y se pregunta: ¿Cuándo existe un polinomio D^*g tal que $(Df|g) = (f|D^*g)$ para todo f? Si tal D^*g existe, se debe tener que

$$(f|D^*g) = f(1)g(1) - f(0)g(0) - (f|Dg)$$

o

$$(f|D^*g + Dg) = f(1)g(1) - f(0)g(0).$$

Con g fijo, $L(f) = f(1)g(1) - f(0)g(0)$ es un funcional lineal del tipo considerado en el Ejemplo 16 y no puede ser de la forma $L(f) = (f|h)$ a menos que $L = 0$. Si D^*g existe, entonces con $h = D^*g + Dg$ se tiene que $L(f) = (f|h)$; por tanto, $g(0) = g(1) = 0$. La existencia de un polinomio apropiado D^*g implica $g(0) = g(1) = 0$. Recíprocamente, si $g(0) = g(1) = 0$, el polinomio $D^*g = -Dg$ satisface $(Df|g) = (f|D^*g)$ para todo f. Si se elige cualquier g para el que $g(0) \neq 0$ o $g(1) \neq 0$, no se puede definir propiamente D^*g, con lo que se concluye que D no tiene adjunto.

Se espera que estos ejemplos aclaren la comprensión de lo que es el adjunto de un operador lineal. Se ve que la operación adjunta, que pasa de T a T^*, se comporta en forma un tanto semejante a la conjugación de números complejos. El siguiente teorema refuerza esta analogía.

Teorema 9. *Sea V un espacio producto interno de dimensión finita. Si T y U son operadores lineales sobre V y c es un escalar*

(i) $(T + U)^* = T^* + U^*$;
(ii) $(cT)^* = \bar{c}T^*$;
(iii) $(TU)^* = U^*T^*$;
(iv) $(T^*)^* = T$.

Demostración. Para demostrar (i), sean α y β vectores en V. Entonces

$$\begin{aligned}((T + U)\alpha|\beta) &= (T\alpha + U\alpha|\beta) \\ &= (T\alpha|\beta) + (U\alpha|\beta) \\ &= (\alpha|T^*\beta) + (\alpha|U^*\beta) \\ &= (\alpha|T^*\beta + U^*\beta) \\ &= (\alpha|(T^* + U^*)\beta).\end{aligned}$$

De la unicidad del adjunto se tiene que $(T + U)^* = T^* + U^*$. Se deja la demostración de (ii) al lector. De las relaciones

$$(TU\alpha|\beta) = (U\alpha|T^*\beta) = (\alpha|U^*T^*\beta)$$

$$(T^*\alpha|\beta) = (\overline{\beta|T^*\alpha}) = (\overline{T\beta|\alpha}) = (\alpha|T\beta). \quad \blacksquare$$

se obtienen (iii) y (iv). ■

El Teorema 9 se enuncia a menudo como sigue: la aplicación $T \to T^*$ es antiisomorfa lineal-conjugada de periodo 2. La analogía con la conjugación compleja, que se ha mencionado anteriormente, está por cierto basada en la observación de que la conjugación compleja tiene las propiedades $\overline{(z_1 + z_2)} = \bar{z}_1 + \bar{z}_2$, $\overline{(z_1 z_2)} = \bar{z}_1 \bar{z}_2$, $\bar{\bar{z}} = z$. Se debe ser cuidadoso y observar la inversión del orden en el producto que impone la operación adjunta: $(UT)^* = T^*U^*$. Cuando se continúe con el estudio de los operadores lineales en un espacio producto interno, se ampliará más esta analogía. En estas líneas algo diremos al respecto. Un número complejo z es real si, y solo si, $z = \bar{z}$. Se podría esperar que el operador lineal T tal que $T = T^*$ se comporte en forma semejante a los números reales. En efecto, es así. Por ejemplo, si T es un operador lineal sobre un espacio *complejo* producto interno de dimensión finita, entonces

(8-15) $$T = U_1 + iU_2$$

donde $U_1 = U_1^*$ y $U_2 = U_2^*$. Así, en cierto sentido, T tiene una «parte real» y una «parte imaginaria». Los operadores U_1 y U_2 que satisfacen $U_1 = U_1^*$ y $U_2 = U_2^*$, y (8-15) son únicos y están dados por

$$U_1 = \frac{1}{2}(T + T^*)$$

$$U_2 = \frac{1}{2i}(T - T^*).$$

Un operador lineal T tal que $T = T^*$ se llama **autoadjunto** (o **hermítico**). Si $\mathcal{B}$ es una base ortonormal de V, entonces

$$[T^*]_{\mathcal{B}} = [T]^*_{\mathcal{B}}$$

y así T es autoadjunto si, y solo si, su matriz en toda base ortonormal es una matriz autoadjunta. Los operadores autoadjuntos son importantes, no solamente porque suministran una especie de parte real y parte imaginaria para el operador general, sino también porque: (1) Los operadores autoadjuntos tienen muchas propiedades especiales. Por ejemplo, para un operador así, existe una base ortonormal de vectores propios. (2) Muchos operadores que surgen en la práctica son autoadjuntos. Las propiedades especiales de los operadores autoadjuntos se considerarán más adelante.

Ejercicios

1. Sea V el espacio C^2 dotado del producto interno canónico. Sea T el operador lineal definido por $T\epsilon_1 = (1, 2)$, $T\epsilon_2 = (i, -1)$. Si $\alpha = (x_1, x_2)$, hallar $T^*\alpha$.

2. Sea T el operador lineal en C^2 definido por $T\epsilon_1 = (1 + i, 2)$, $T\epsilon_2 = (i, i)$. Usando el producto interno canónico, hallar la matriz de T^* en la base ordenada canónica. ¿Conmuta T con T^*?

3. Sea $V = C^3$ con el producto interno canónico. Sea T el operador lineal sobre V cuya matriz en la base ordenada canónica está definida por

$$A_{jk} = i^{j+k}, \qquad (i^2 = -1).$$

Hallar una base para el espacio nulo de T^*.

4. Sean V un espacio producto interno de dimensión finita y T un operador lineal sobre V. Demostrar que la imagen por T^* es el complemento ortogonal del espacio nulo de T.

5. Sean V un espacio producto interno de dimensión finita y T un operador lineal sobre V. Si T es inversible, demostrar que T^* es inversible y que $(T^*)^{-1} = (T^{-1})^*$.

6. Sean V un espacio producto interno y β, γ vectores dados de V. Demostrar que $T\alpha = (\alpha|\beta)\gamma$ define un operador lineal sobre V. Demostrar que T tiene un adjunto y dar explícitamente T^*.

Supóngase ahora que $V = C^n$ con el producto interno canónico $\beta = (y_1, \ldots, y_n)$ y $\gamma = (x_1, \ldots, x_n)$. ¿Cuál es el elemento j, k de la matriz de T en la base ordenada canónica? ¿Cuál es el rango de esta matriz?

7. Demostrar que el producto de dos operadores autoadjuntos es autoadjunto si, y solo si, los dos operadores conmutan.

8. Sea V el espacio vectorial de los polinomios sobre R de grado menor o igual que 3 con el producto interno

$$(f|g) = \int_0^1 f(t)g(t)\,dt.$$

Si t es un número real, hallar el polinomio g_t de V tal que $(f|g_t) = f(t)$ para todo f de V.

9. Sea V el espacio producto interno del Ejercicio 8 y sea D el operador derivación sobre V. Hallar D^*.

10. Sea V el espacio de las matrices $n \times n$ sobre los números complejos con el producto interno $(A|B) = \operatorname{tr}(AB^*)$. Sea P una matriz dada en V, inversible, y sea T_P el operador lineal sobre V definido por $T_P(A) = P^{-1}AP$. Hallar el adjunto de T_P.

11. Sea V un espacio producto interno de dimensión finita y sea E un operador lineal idempotente sobre V; es decir, $E^2 = E$. Demostrar que E es autoadjunto si, y solo si, $EE^* = E^*E$.

12. Sea V un espacio producto interno *complejo* de dimensión finita y sea T un operador lineal sobre V. Demostrar que T es autoadjunto si, y solo si, $(T\alpha|\alpha)$ es real para todo α de V.

8.4. Operadores unitarios

En esta sección se considerará el concepto de un isomorfismo entre dos espacios producto interno. Si V y W son espacios vectoriales, un isomorfismo de V sobre W es una transformación lineal inyectiva de V sobre W; es decir, una correspondencia biunívoca entre los elementos de V y de W que «preservan» las operaciones de espacio vectorial. Ahora bien, un espacio producto interno consiste en un espacio vectorial y un producto interno definido sobre dicho espacio. Así, pues, cuando V y W son espacios producto interno, se exigirá de un isomorfismo de V sobre W no solo que preserve las operaciones lineales, sino que también preserve el producto interno. Un isomorfismo de un espacio producto interno sobre sí mismo se llama «operador unitario» sobre ese espacio. Se considerarán varios ejemplos de operadores unitarios y se establecerán sus propiedades básicas.

Definición. *Sean V y W espacios producto interno sobre el mismo cuerpo y sea T una transformación lineal de V en W. Se dice que T* **preserva productos internos** *si* $(T\alpha|T\beta) = (\alpha|\beta)$ *para todo* α, β *de V. Un* **isomorfismo** *de V sobre W es un isomorfismo T de espacio vectorial de V sobre W que también preserva productos internos.*

Si T preserva productos internos, entonces $\|T\alpha\| = \|\alpha\|$ y así T es necesariamente no singular. Así que un isomorfismo de V sobre W puede ser definido también como una transformación lineal de V sobre W que preserva productos internos. Si T es un isomorfismo de V sobre W, entonces T^{-1} es un isomorfismo de W sobre V; luego, cuando tal T existe, se dirá simplemente que V y W son **isomorfos.** Naturalmente, el isomorfo de espacio producto interno es una relación de equivalencia.

Teorema 10. *Sean V y W espacios producto interno de dimensión finita sobre el mismo cuerpo y que tienen la misma dimensión. Si T es una transformación lineal de V en W, las siguientes afirmaciones son equivalentes.*

(i) *T preserva los productos internos.*
(ii) *T es un isomorfismo (en un espacio producto interno).*
(iii) *T aplica toda base ortonormal de V sobre una base ortonormal de W.*
(iv) *T aplica cierta base ortonormal de V sobre una base ortonormal de W.*

Demostración. (i) $\longrightarrow$ (ii). Si T preserva los productos internos, entonces $\|T\alpha\| = \|\alpha\|$ para todo α de V. Así, T es no singular, y como $\dim V = \dim W$, se sabe que T es un isomorfismo de espacio vectorial.

(ii) $\longrightarrow$ (iii). Supóngase que T sea un isomorfismo. Sea $\{\alpha_1, \ldots, \alpha_n\}$ una base ortonormal de V. Como T es un isomorfismo de espacio vectorial y $\dim W = \dim V$, se sigue que $\{T\alpha_1, \ldots, T\alpha_n\}$ es una base de W. Como T también preserva los productos internos, $(T\alpha_j|T\alpha_k) = (\alpha_j|\alpha_k) = \delta_{jk}$.

(iii) $\longrightarrow$ (iv). Esta no requiere comentario.

(iv) $\longrightarrow$ (i). Sea $\{\alpha_1, \ldots, \alpha_n\}$ una base ortonormal de V tal que $\{T\alpha_1, \ldots, T\alpha_n\}$ sea una base ortonormal de W. Entonces

$$(T\alpha_j|T\alpha_k) = (\alpha_j|\alpha_k) = \delta_{jk}.$$

Para cualquier $\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n$ y $\beta = y_1\alpha_1 + \cdots + y_n\alpha_n$ de V, tenemos

$$\begin{aligned}(\alpha|\beta) &= \sum_{j=1}^{n} x_j\bar{y}_j \\ (T\alpha|T\beta) &= \Big(\sum_j x_jT\alpha_j\Big|\sum_k y_kT\alpha_k\Big) \\ &= \sum_j\sum_k x_j\bar{y}_k(T\alpha_j|T\alpha_k) \\ &= \sum_{j=1}^{n} x_j\bar{y}_j\end{aligned}$$

y así, pues, T preserva los productos internos. ∎

Corolario. *Sean V y W espacios producto interno de dimensión finita sobre el mismo cuerpo. Entonces V y W son isomorfos si, y solo si, tienen la misma dimensión.*

Demostración. Si $\{\alpha_1, \ldots, \alpha_n\}$ es una base ortonormal de V y $\{\beta_1, \ldots, \beta_n\}$ es una base ortonormal de W, sea T la transformación lineal de V en W definida por $T\alpha_j = \beta_j$. Entonces T es un isomorfismo de V sobre W. ▌

Ejemplo 22. Si V es un espacio producto interno de dimensión finita, entonces cada base ordenada ortonormal $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ determina un isomorfismo de V sobre F^n con el producto interno canónico. El isomorfismo no es más que

$$T(x_1\alpha_1 + \cdots + x_n\alpha_n) = (x_1, \ldots, x_n).$$

Existe el isomorfismo, aparentemente distinto, que $\mathcal{B}$ determina de V sobre el espacio $F^{n\times 1}$ con $(X|Y) = Y^*X$ como producto interno. El isomorfismo es

$$\alpha \to [\alpha]_{\mathcal{B}}$$

es decir, la transformación que aplica α en su matriz de coordenadas en la base ordenada $\mathcal{B}$. Para cualquier base $\mathcal{B}$, éste es un isomorfismo del espacio vectorial; sin embargo, es un isomorfismo de los dos espacios con producto interno si, y solo si, es ortogonal.

Ejemplo 23. Aquí se presenta un isomorfismo algo menos superficial. Sea W el espacio de todas las matrices 3×3, A, sobre R antisimétricas; es decir, $A^t = -A$. Se dota a W del producto interno $(A|B) = \frac{1}{2}\operatorname{tr}(AB^t)$; el $\frac{1}{2}$ es por mera conveniencia. Sea V el espacio R^3 con el producto interno canónico. Sea T una transformación lineal de V en W definida por

$$T(x_1, x_2, x_3) = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}.$$

Entonces T aplica V sobre W, y haciendo

$$A = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{bmatrix}$$

tenemos

$$\begin{aligned} \operatorname{tr}(AB^t) &= x_3y_3 + x_2y_2 + x_3y_3 + x_2y_2 + x_1y_1 \\ &= 2(x_1y_1 + x_2y_2 + x_3y_3). \end{aligned}$$

Así, pues, $(\alpha|\beta) = (T\alpha|T\beta)$ y T es un isomorfismo de espacio vectorial. Obsérvese que T aplica la base canónica $\{\epsilon_1, \epsilon_2, \epsilon_3\}$ sobre la base ortonormal que consta de las tres matrices

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Ejemplo 24. No siempre es particularmente conveniente describir un isomorfismo en términos de las bases ortonormales. Por ejemplo, supóngase que $G = P^*P$, donde P es una matriz $n \times n$, inversible, con elementos complejos. Sea V el espacio de las matrices complejas $n \times 1$ con el producto interno $[X|Y] = Y^*GX$. Sea W el mismo espacio vectorial, con el producto interno canónico $(X|Y) = Y^*X$. Se sabe que V y W son espacios producto interno isomorfos. Parecería que el modo más conveniente de dar un isomorfismo entre V y W es el siguiente: sea T la transformación lineal de V en W definida por $T(X) = PX$. Entonces

$$\begin{aligned}(TX|TY) &= (PX|PY)\\ &= (PY)^*(PX)\\ &= Y^*P^*PX\\ &= Y^*GX\\ &= [X|Y].\end{aligned}$$

Luego T es un isomorfismo.

Ejemplo 25. Sea V el espacio de todas las funciones continuas de valor real en el intervalo unitario $0 \leq t \leq 1$, con el producto interno

$$[f|g] = \int_0^1 f(t)g(t)t^2\,dt.$$

Sea W el mismo espacio vectorial con el producto interno

$$(f|g) = \int_0^1 f(t)g(t)\,dt.$$

Sea T la transformación lineal de V en W dada por

$$(Tf)(t) = tf(t).$$

Entonces $(Tf|Tg) = [f|g]$, y así T preserva los productos internos; sin embargo, T *no* es un isomorfismo de V sobre W, ya que la imagen de T no es todo W. Por supuesto, esto sucede porque el espacio vectorial básico no es de dimensión finita.

Teorema 11. *Sean V y W espacios producto interno sobre el mismo cuerpo y sea T una transformación lineal de V en W. Entonces T preserva productos internos si, y solo si, $\|T\alpha\| = \|\alpha\|$ para todo α en V.*

Demostración. Si T preserva productos internos, entonces T «preserva normas». Supóngase que $\|T\alpha\| = \|\alpha\|$ para todo α de V. Entonces $\|T\alpha\|^2 = \|\alpha\|^2$. Ahora, usando la identidad de polarización apropiada, (8-3) o (8-4), y el hecho de que T es lineal, se obtiene fácilmente que $(\alpha|\beta) = (T\alpha|T\beta)$ para todo α, β en V ▮

Definición. *Un* **operador unitario** *en un espacio producto interno es un isomorfismo del espacio sobre sí mismo.*

El producto de dos operadores unitarios es unitario. En efecto, si U_1 y U_2

son unitarios, entonces U_2U_1 es inversible y $\|U_2U_1\alpha\| = \|U_1\alpha\| = \|\alpha\|$ para cada α. También el inverso del operador unitario es unitario, ya que $\|U\alpha\| = \|\alpha\|$, dice que $\|U^{-1}\beta\| = \|\beta\|$, donde $\beta = U\alpha$. Dado que el operador unitario es evidentemente unitario, se ve que el conjunto de todos los operadores unitarios en un espacio producto interno es un grupo para la operación de composición.

Si V es un espacio producto interno y U es un operador lineal sobre V, el Teorema 10 dice que U es unitario si, y solo si, $(U\alpha|U\beta) = (\alpha|\beta)$ para todo α, β de V; o si, y solo si, para cierta (toda) base ortonormal $\{\alpha_1, \ldots, \alpha_n\}$ es cierto que $\{U\alpha_1, \ldots, U\alpha_n\}$ es una base ortonormal.

Teorema 12. *Sea U un operador lineal sobre un espacio producto interno V. Entonces U es unitario si, y solo si, el adjunto U^* de U existe y $UU^* = U^*U = I$.*

Demostración. Supóngase que U es unitario. Entonces U es inversible y

$$(U\alpha|\beta) = (U\alpha|UU^{-1}\beta) = (\alpha|U^{-1}\beta)$$

para todo α, β. Luego U^{-1} es el adjunto de U.

Recíprocamente, supóngase que existe U^* y que $UU^* = U^*U = I$. Entonces U es inversible, con $U^{-1} = U^*$. De modo que solo se necesita demostrar que U preserva productos internos. Tenemos que

$$\begin{aligned}(U\alpha|U\beta) &= (\alpha|U^*U\beta)\\ &= (\alpha|I\beta)\\ &= (\alpha|\beta)\end{aligned}$$

para todo α, β. ∎

Ejemplo 26. Consideremos $C^{n\times 1}$ con el producto interno $(X|Y) = Y^*X$. Sea A una matriz $n \times n$ sobre C, y sea U el operador lineal definido por $U(X) = AX$. Entonces

$$(UX|UY) = (AX|AY) = Y^*A^*AX$$

para todo X, Y. Luego U es unitario si, y solo si, $A^*A = I$.

Definición. *Una matriz compleja $n \times n$, A, se llama* **unitaria** *si $A^*A = I$.*

Teorema 13. *Sea V un espacio producto interno de dimensión finita y sea U un operador lineal sobre V. Entonces U es unitario si, y solo si, la matriz de U en alguna (o toda) base ordenada ortonormal es una matriz unitaria.*

Demostración. A esta altura, esto casi no es un teorema, y se formuló más que todo por insistir. Si $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ es una base ordenada ortonormal de V y A es la matriz de U respecto de $\mathcal{B}$, entonces $A^*A = I$ si, y solo si, $U^*U = I$. El resultado se desprende ahora del Teorema 12. ∎

Sea A una matriz $n \times n$. La afirmación de que A es unitaria solo quiere decir que

$$(A^*A)_{jk} = \delta_{jk}$$

o

$$\sum_{r=1}^{n} \overline{A_{rj}} A_{rk} = \delta_{jk}.$$

En otras palabras, esto significa que las columnas de A forman un conjunto ortonormal de matrices columna con respecto al producto interno canónico $(X/Y) = Y^*X$. Como $A^*A = I$ si, y solo si, $AA^* = I$, vemos que A es unitaria exactamente cuando las filas de A constituyen un conjunto ortonormal de n-tuples de C_n (con el producto interno canónico). Así, usando los productos internos canónicos, A es unitaria si, y solo si, las filas y las columnas de A son conjuntos ortonormales. Se ve aquí un ejemplo del poder del teorema que establece que una inversa latèral (a un lado) para una matriz es inversa a ambos lados. Aplicando este teorema como se hizo anteriormente, por ejemplo, a matrices reales, tenemos lo siguiente: supóngase que tenemos un arreglo en cuadro de números reales tales que la suma de los cuadrados de los elementos de cada fila es 1 y las filas distintas son ortogonales. Entonces, la suma de los cuadrados de los elementos de cada columna es 1 y las columnas distintas son ortogonales. Escríbase la demostración de esto para un arreglo de 3 × 3 sin usar lo que se conoce de las matrices y se quedará bastante impresionado.

Definición. *Una matriz $n \times n$ real o compleja, A, se dice* **ortogonal** *si $A^tA = I$*

Una matriz ortogonal real es unitaria, y una matriz unitaria es ortogonal si, y solo si, cada uno de sus elementos es real.

Ejemplo 27. Damos algunos ejemplos de matrices unitarias y ortogonales.

(a) A es la matriz 1×1, c, es ortogonal si, y solo si, $c = \pm 1$, y es unitaria si, y solo si, $\bar{c}c = 1$. La última condición quiere decir (naturalmente) que $|c| = 1$, o $c = e^{i\theta}$, donde θ es real.

(b) Sea

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Entonces A es ortogonal si, y solo si,

$$A^t = A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

El determinante de una matriz ortogonal es, como fácilmente se ve, ± 1. Así A es ortogonal si, y solo si,

$$A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

o

$$A = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$$

donde $a^2 + b^2 = 1$. Los dos casos se distinguen por el valor de det A.

(c) Las relaciones, bien conocidas, entre funciones trigonométricas muestran que la matriz

$$A_\theta = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$$

es ortogonal. Si θ es un número real, entonces A_θ es la matriz en la base ordenada canónica de R^2 del operador lineal U_θ, rotación de ángulo θ. La afirmación de que A_θ es una matriz ortogonal real (por tanto unitaria) solo quiere decir que U_θ es un operador unitario, es decir, preserva productos escalares.

(d) Sea

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Entonces A es unitaria si, y solo si,

$$\begin{bmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

El determinante de una matriz unitaria tiene valor absoluto 1 y es, por tanto, un número complejo de la forma $e^{i\theta}$, θ real. Así A es unitaria si, y solo si,

$$A = \begin{bmatrix} a & b \\ -e^{i\theta}\bar{b} & e^{i\theta}\bar{a} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix}\begin{bmatrix} a & b \\ -\bar{b} & \bar{a} \end{bmatrix}$$

donde θ es un número real, y a, b son números complejos tales que $|a|^2 + |b|^2 = 1$.

Como se observó antes, los operadores unitarios sobre un espacio producto interno forman un grupo. De esto y del Teorema 13 se sigue que el conjunto $U(n)$ de todas las matrices unitarias $n \times n$ es también un grupo. Así la inversa de una matriz unitaria y el producto de dos matrices unitarias son también matrices unitarias. Por cierto, esto es fácil verlo directamente. Una matriz $n \times n$, A, con elementos complejos es unitaria si, y solo si, $A^{-1} = A^*$. Así, si A es unitaria, tenemos que $(A^{-1})^{-1} = A = (A^*)^{-1} = (A^{-1})^*$. Si A y B son matrices unitarias $n \times n$, entonces $(AB)^{-1} = B^{-1}A^{-1} = B^*A^* = (AB)^*$.

El proceso de Gram-Schmidt en C^n tiene un corolario interesante para las matrices en que entra el grupo $U(n)$.

Teorema 14. *Para toda matriz compleja $n \times n$, inversible, B, existe una única matriz triangular inferior M, con elementos positivos en la diagonal principal, de modo que MB es unitaria.*

Demostración. Las filas $\beta_1, \ldots, B_n$ de B forman una base de C^n. Sean $\alpha_1, \ldots, \alpha_n$ los vectores que se obtienen de $\beta_1, \ldots, \beta_n$ por el proceso de Gram-Schmidt. Entonces, para $1 \leq k \leq n$, $\{\alpha_1, \ldots, \alpha_k\}$ es una base ortogonal del subespacio generado por $\{\beta_1, \ldots, \beta_k\}$, y

$$\alpha_k = \beta_k - \sum_{j<k} \frac{(\beta_k|\alpha_j)}{||\alpha_j||^2}\alpha_j.$$

Luego, para todo k, existen escalares únicos C_{kj} tales que

$$\alpha_k = \beta_k - \sum_{j<k} C_{kj}\beta_j.$$

Sea U la matriz unitaria con filas

$$\frac{\alpha_1}{||\alpha_1||}, \dots, \frac{\alpha_n}{||\alpha_n||}$$

y M la matriz definida por

$$M_{kj} = \begin{cases} -\frac{1}{||\alpha_k||} \cdot C_{kj}, & \text{si } j < k \\ \frac{1}{||\alpha_k||}, & \text{si } j = k \\ 0, & \text{si } j > k. \end{cases}$$

Entonces M es triangular inferior, ya que sus elementos sobre la diagonal principal son 0. Los elementos M_{kk} de M, de la diagonal principal, son todos >0, y

$$\frac{\alpha_k}{||\alpha_k||} = \sum_{j=1}^{n} M_{kj}\beta_j, \qquad 1 \le k \le n.$$

Ahora estas igualdades simplemente dicen que

$$U = MB.$$

Para demostrar la unicidad de M, sea $T^+(n)$ el conjunto de todas las matrices complejas $n \times n$ triangulares, inferiores con elementos positivos en la diagonal principal. Supóngase que M_1 y M_2 sean elementos de $M^+(n)$ tales que M_iB está en $U(n)$ para $i = 1, 2$. Entonces, por ser $U(n)$ un grupo

$$(M_1B)(M_2B)^{-1} = M_1M_2^{-1}$$

está en $U(n)$. Por otro lado, aun cuando no es enteramente obvio, $T^+(n)$ es también un grupo para la multiplicación matricial. Un modo de verlo es considerar las propiedades geométricas de las transformaciones lineales

$$X \to MX, \qquad (M \text{ en } T^+(n))$$

sobre el espacio de las matrices columnas. Así M_2^{-1}, $M_1M_2^{-1}$ y $(M_1M_2^{-1})^{-1}$ están todas en $T^+(n)$. Pero, como $M_1M_2^{-1}$ está en $U(n)$, $(M_1M_2^{-1})^{-1} = (M_1M_2^{-1})^*$. La transpuesta o la conjugada transpuesta de cualquier matriz triangular inferior es una matriz triangular superior. Por tanto, $M_1M_2^{-1}$ es simultáneamente triangular superior e inferior, es decir, diagonal. Una matriz diagonal es unitaria si, y solo si, cada uno de sus elementos de la diagonal principal tiene valor absoluto 1; si los elementos de la diagonal son todos positivos, deben ser iguales a 1. Luego $M_1M_2^{-1} = I$, con lo que $M_1 = M_2$. ∎

Sea $GL(n)$ el conjunto de todas las matrices complejas $n \times n$ e inversibles. Entonces $GL(n)$ es también un grupo para la multiplicación matricial. Este grupo se llama el **grupo lineal general.** El Teorema 14 es equivalente al siguiente resultado.

Corolario. *Para toda B de $GL(n)$, existen matrices únicas N y U tales que N está en $T^+(n)$, U en $U(n)$, y*

$$B = N \cdot U.$$

Demostración. Por el teorema existe una matriz única M en $T^+(n)$ tal que MB está en $U(n)$. Sea $MB = U$ y $N = M^{-1}$. Entonces N está en $T^+(n)$ y $B = N \cdot U$. Por otro lado, si se dan elementos N y U cualesquiera tales que N está en $T^+(n)$, U está en $U(n)$ y $B = N \cdot U$, entonces $N^{-1}B$ está en $U(n)$ y N^{-1} es la única matriz M que está caracterizada por el teorema. Además U es necesariamente $N^{-1}B$. ∎

Ejemplo 28. Sean x_1 y x_2 números reales tales que $x_1^2 + x_2^2 = 1$ y $x_1 \neq 0$. Sea

$$B = \begin{bmatrix} x_1 & x_2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Aplicando el proceso de Gram-Schmidt a las filas de B, se obtienen los vectores

$$\begin{aligned} \alpha_1 &= (x_1, x_2, 0) \\ \alpha_2 &= (0, 1, 0) - x_2(x_1, x_2, 0) \\ &= x_1(-x_2, x_1, 0) \\ \alpha_3 &= (0, 0, 1). \end{aligned}$$

Sea U la matriz con filas α_1, (α_2/x_1), α_3. Entonces U es unitaria, y

$$U = \begin{bmatrix} x_1 & x_2 & 0 \\ -x_2 & x_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\dfrac{x_2}{x_1} & \dfrac{1}{x_1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Ahora, multiplicando por la inversa de

$$M = \begin{bmatrix} 1 & 0 & 0 \\ -\dfrac{x_2}{x_1} & \dfrac{1}{x_1} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

se tiene que

$$\begin{bmatrix} x_1 & x_2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ x_2 & x_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & 0 \\ -x_2 & x_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Consideremos ahora brevemente el cambio de coordenadas en un espacio producto interno. Supóngase que V es un espacio producto interno de dimensión finita, y que $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ y $\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ son dos bases orde-

nadas *ortonormales* de V. Existe una única matriz $n \times n$ (necesariamente inversible), P, tal que

$$[\alpha]_{\mathcal{B}'} = P^{-1}[\alpha]_{\mathcal{B}}$$

para todo α en V. Si U es el operador lineal único sobre V definido por $U\alpha_j = \alpha'_j$, entonces P es la matriz de U en la base ordenada $\mathcal{B}$;

$$\alpha'_k = \sum_{j=1}^{n} P_{jk}\alpha_j.$$

Como $\mathcal{B}$ y $\mathcal{B}'$ son bases ortonormales, U es un operador unitario y P es una matriz unitaria. Si T es un operador lineal cualquiera sobre V, entonces

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P = P^*[T]_{\mathcal{B}}P.$$

Definición. *Sean A y B matrices complejas $n \times n$. Se dice que B es* **unitariamente equivalente a** *A si existe una matriz unitaria $n \times n$, P, tal que $B = P^{-1}AP$. Se dice que B es* **ortogonalmente equivalente a** *A si existe una matriz ortogonal $n \times n$, P, tal que $B = P^{-1}AP$.*

Con esta definición, lo que se observó anteriormente puede enunciarse como sigue: si $\mathcal{B}$ y $\mathcal{B}'$ son dos bases ordenadas ortonormales para V, entonces, para todo operador lineal T sobre V, la matriz $[T]_{\mathcal{B}}$ es unitariamente equivalente a la matriz $[T]_{\mathcal{B}}$. En el caso que V sea un espacio producto interno real, estas matrices son ortogonalmente equivalentes a través de una matriz ortogonal real.

Ejercicios

1. Hallar una matriz unitaria que no sea ortogonal y encontrar una matriz ortogonal que no sea unitaria.

2. Sea V el espacio de las matrices complejas $n \times n$ con producto interno $(A|B) = \text{tr}(AB^*)$. Para cada M en V, sea T_M el operador lineal definido por $T_M(A) = MA$. Demostrar que T_M es unitario si, y solo si, M es una matriz unitaria.

3. Sea V el conjunto de los números complejos considerados como espacio vectorial real.

(a) Demostrar que $(\alpha|\beta) = \text{Re}(\alpha\bar{\beta})$ define un producto interno sobre V.

(b) Mostrar un isomorfismo (de espacio producto interno) de V sobre R^2 con el producto interno canónico.

(c) Para cada γ de V, sea M_γ el operador lineal sobre V definido por $M_\gamma(\alpha) = \gamma\alpha$. Demostrar que $(M_\gamma)^* = M_{\bar{\gamma}}$.

(d) ¿Para qué números complejos γ es M_γ autoadjunto?

(e) ¿Para cuáles γ es M_γ unitario?

(f) ¿Para cuáles γ es M_γ positivo?

(g) ¿Cuál es det (M_γ)?

(h) Hallar la matriz de M_γ en la base $\{1, i\}$.

(i) Si T es un operador lineal en V, hallar condiciones necesarias y suficientes para T para que sea un M_γ.

(j) Encontrar un operador unitario sobre V que no sea un M_γ.

4. Sea $V = R^2$ con el producto interno canónico. Si U es un operador unitario sobre V, demostrar que la matriz de U en la base ordenada canónica es

$$\begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} \quad \text{o} \quad \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ \operatorname{sen}\theta & -\cos\theta \end{bmatrix}$$

para algún real θ, $0 \leq \theta \leq 2$. Sea U_θ el operador lineal que corresponde a la primera matriz; es decir, U_θ es una rotación de ángulo θ. Ahora observar que todo operador unitario sobre V es una rotación o una simetría en torno al eje ≤ 1 seguida de una rotación.

(a) ¿Qué es $U_\theta U_\phi$?

(b) Demostrar que $U_\theta^* = U_\theta$.

(c) Sea ϕ un número real fijo y sea $\mathcal{B} = \{\alpha_1, \alpha_2\}$ la base ortonornal obtenida por rotación de $\{\epsilon_1, \epsilon_2\}$ en el ángulo ϕ; es decir, $\alpha_j = U_\phi \epsilon_j$. Si θ es otro número real, ¿cuál es la matriz de U_9 en la base ordenada $\mathcal{B}$?

5. Sea $V = R^3$ con el producto interno canónico. Sea W el plano generado por $\alpha = (1, 1, 1)$ y $\beta = (1, 1, -2)$. Sea U el operador lineal definido geométricamente como sigue: U es la rotación de ángulo θ en torno a la recta que pasa por el origen ortogonal a W. En realidad hay dos de estas rotaciones —se elige una. Hallar la matriz de U en la base ordenada canónica. (He aquí un camino que se podría seguir: hallar que α_1 y α_2 forman una base ortonormal de W. Sea α_3 un vector de norma 1 que es ortogonal a W. Hallar la matriz de U en la base $\{\alpha_1, \alpha_2, \alpha_3\}$. Hágase un cambio de base.)

6. Sea V un espacio producto interno de dimensión finita y sea W un subespacio de V. Entonces $V = W \oplus W^\perp$, esto es, todo α de V se expresa unívocamente en la forma $\alpha = \beta + \gamma$, con β en W y γ en $W^\perp$. Se define un operador lineal U por $U\alpha = \beta - \gamma$.

(a) Demostrar que U es autoadjunto y unitario.

(b) Si V es R^3, con el producto interno canónico, y W es el subespacio generado por $(1, 0, 1)$, hallar la matriz de U en la base ordenada canónica.

7. Sea V un espacio *complejo* con producto interno y T un operador lineal *autoadjunto* sobre V. Demostrar que

(a) $\|\alpha + iT\alpha\| = \|\alpha - iT\alpha\|$ para todo α en V.

(b) $\alpha + iT\alpha = \beta + iT\beta$ si, y solo si, $\alpha = \beta$.

(c) $I + iT$ es no singular.

(d) $I - iT$ es no singular.

(e) Supongamos ahora que V es de dimensión finita y demostremos que

$$U = (I - iT)(I + iT)^{-1}$$

es un operador unitario; U se llama la **transformación de Cayley** de T. En cierto sentido $U = f(T)$, donde $f(x) = (1 - ix)/(1 + ix)$.

8. Si θ es un número real, demostrar que las siguientes matrices son equivalentes unitariamente

$$\begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}, \quad \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}.$$

9. Sean V un espacio producto interno de dimensión finita y T un operador lineal positivo sobre V. Sea p_T el producto interno sobre V definido por $p_T(\alpha, \beta) = (T\alpha, \beta)$. Sea U un operador lineal en V y U^* su adjunto con respecto a $(\ |\ )$. Demostrar que U es unitario con respecto al producto interno p_T si, y solo si, $T = U^*TU$.

10. Sea V un espacio producto interno de dimensión finita. Para cada α, β en V, sea $T_{\alpha,\beta}$ el operador lineal en V definido por $T_{\alpha,\beta}(\gamma) = (\gamma|\beta)\alpha$. Demostrar que

(a) $T^*_{\alpha,\beta} = T_{\beta,\alpha}$.
(b) traza $(T_{\alpha,\beta}) = (\alpha|\beta)$.
(c) $T_{\alpha,\beta}T_{\gamma,\delta} = T_{\alpha,(\beta|\gamma)\delta}$.
(d) ¿En qué condiciones es $T_{\alpha,\beta}$ autoadjunto?

11. Sea V un espacio producto interno de dimensión finita sobre el cuerpo F y sea $L(V, V)$ el espacio de los operadores lineales sobre V. Demostrar que existe un producto interno único sobre $L(V, V)$ con la propiedad de que $\|T_{\alpha,\beta}\|^2 = \|\alpha\|^2\|\beta\|^2$ para todo α, β de V ($T_{\alpha,\beta}$ es el operador definido en el Ejercicio 10). Hallar un isomorfismo entre $L(V, V)$, con este producto interno, y el espacio de las matrices $n \times n$ sobre F, con el producto interno $(A|B) = \text{tr}\,(AB^*)$.

12. Sea V un espacio producto interno de dimensión finita. En el Ejercicio 6 mostramos como construir operadores lineales sobre V que son autoadjuntos y unitarios a la vez. Demostrar ahora que no hay otros, es decir, que todo operador autoadjunto y unitario surge de cierto subespacio W como se describió en el Ejercicio 6.

13. Sean V y W espacios producto interno de dimensión finita que tienen la misma dimensión. Sea U un isomorfismo de V sobre W. Demostrar que

(a) La aplicación $T \rightarrow UTU^{-1}$ es un isomorfismo del espacio vectorial $L(V, V)$ sobre el espacio vectorial $L(W, W)$.
(b) traza (UTU^{-1}) = traza (T) para todo T de $L(V, V)$.
(c) $UT_{\alpha,\beta}U^{-1} = T_{U\alpha, U\beta}$ ($T_{\alpha,\beta}$ definido en el Ejercicio 10).
(d) $(UTU^{-1})^* = UT^*U^{-1}$
(e) Si se dota $L(V, V)$ del producto interno $(T_1|T_2)$ = traza $(T_1T_2^*)$, y análogamente para $L(W, W)$, entonces $T \rightarrow UTU^{-1}$ es un isomorfismo de espacios producto interno.

14. Si V es un espacio producto interno, un **movimiento rígido** es cierta función T de V en V (no necesariamente lineal) tal que $\|T\alpha - T\beta\| = \|\alpha - \beta\|$ para todo α, β de V. Ejemplo de movimiento rígido es un operador lineal unitario. Otro ejemplo es la traslación de vector dado γ:

$$T_\gamma(\alpha) = \alpha + \gamma$$

(a) Sea $V = R^2$, con el producto interno canónico. Supóngase que T es un movimiento rígido de V y que $T(0) = 0$. Demostrar que T es operador lineal y unitario.
(b) Usar el resultado de la parte (a) para demostrar que todo movimiento rígido de R^2 está compuesto de una traslación, seguida de un operador unitario.
(c) Demostrar ahora que un movimiento rígido de R^2 es una traslación seguida de una rotación, o una traslación seguida de una simetría seguida de una rotación.

15. Un operador unitario sobre R^4 (dotado del producto interno canónico) es simplemente un operador lineal que preserva la forma cuadrática,

$$\|(x, y, z, t)\|^2 = x^2 + y^2 + z^2 + t^2$$

esto es, un operador lineal U tal que $\|U\alpha\|^2 = \|\alpha\|^2$ para todo α de R^4. En cierta parte de la teoría de la relatividad es de interés hallar los operadores lineales T que preservan la forma

$$\|(x, y, z, t)\|_L^2 = t^2 - x^2 - y^2 - z^2.$$

Pero aquí $\|\ \ \|_L^2$ no proviene de un producto interno, sino de algo llamado la «métrica de

Lorentz» (en la que no entraremos). Por esa razón un operador lineal T sobre R^4 tal que $\|T\alpha\|_L^2 = \|\alpha\|_L^2$, para todo α de R^4, se llama **transformación de Lorentz.**

(a) Demostrar que la función U definida por

$$U(x, y, z, t) = \begin{bmatrix} t + x & y + iz \\ y - iz & t - x \end{bmatrix}$$

es un isomorfismo de R^4 sobre el espacio vectorial real H de las matrices complejas 2×2 autoadjuntas.

(b) Demostrar que $\|\alpha\|_L^2 = \det(U\alpha)$.

(c) Supóngase que T es un operador lineal (real) sobre el espacio H de las matrices 2×2 autoadjuntas. Demostrar que $L = U^{-1}TU$ es un operador lineal sobre R^4.

(d) Sea M una matriz compleja 2×2 cualquiera. Mostrar que $T_M(A) = M^*AM$ define un operador lineal T_M sobre H (asegurarse de que T_M aplica H en H).

(e) Si M es una matriz 2×2 tal que $|\det M| = 1$, demostrar que $L_M = U^{-1}T_MU$ es una transformación de Lorentz en R^4.

(f) Hallar una transformación de Lorentz que no sea un L_M.

8.5. Operadores normales

El objetivo principal de esta sección es la solución del siguiente problema. Si T es un operador lineal sobre un espacio producto interno de dimensión finita V, ¿en qué condiciones tiene V una base ortonormal formada por vectores propios de T? En otras palabras, ¿cuándo existe una base *ortonormal* $\mathcal{B}$ de V, tal que la matriz de T en la base $\mathcal{B}$ sea diagonal?

Comenzaremos deduciendo unas condiciones necesarias para T, que más adelante se verá que son suficientes. Supongamos que $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ es una base ortonormal de V con la propiedad

(8-16) $$T\alpha_j = c_j\alpha_j, \qquad j = 1, \ldots, n.$$

Esto solo dice que la matriz de T en la base ordenada $\mathcal{B}$ es la matriz diagonal con elementos $c_1, \ldots, c_n$ en la diagonal. El operador adjunto T^* está representado en esta misma base ordenada por la matriz transpuesta conjugada; es decir, la matriz diagonal con elementos $\bar{c}_1, \ldots, \bar{c}_n$ en la diagonal. Si V es un espacio producto interno real, los escalares $c_1, \ldots, c_n$ son (claro está) reales y, por tanto, debe tenerse que $T = T^*$. En otras palabras, si V es un espacio producto interno real de dimensión finita y T es un operador lineal para el que existe una base ortonormal de vectores propios, entonces T debe ser autoadjunto. Si V es un espacio producto interno complejo, los escalares $c_1, \ldots, c_n$ no necesitan ser reales; es decir, T no necesita ser autoadjunto. Pero obsérvese que T debe satisfacer

(8-17) $$TT^* = T^*T.$$

En efecto, dos matrices diagonales cualesquiera conmutan, y como T y T^* están ambos representados por matrices diagonales en la base ordenada $\mathcal{B}$, se tiene (8-17). Es más bien notable que, en el caso complejo, esta condición sea también suficiente para implicar la existencia de una base ortonormal de vectores propios.

Definición. *Sean V un espacio producto interno de dimensión finita y T un operador lineal sobre V. Se dice que T es* **normal** *si conmuta con su adjunto; es decir,* $TT^* = T^*T$

Todo operador autoadjunto es normal, como también todo operador unitario; sin embargo, las sumas y productos de operadores normales no son en general normales. Aunque no es necesario, comenzaremos el estudio de los operadores normales considerando operadores autoadjuntos.

Teorema 15. *Sean V un espacio producto interno y T un operador lineal autoadjunto sobre V. Entonces todo valor propio de T es real y los vectores propios de T, asociados a valores propios distintos, son ortogonales.*

Demostración. Supóngase que c es un valor propio de T; es decir, que $T\alpha = c\alpha$ para algún vector no nulo α. Entonces

$$\begin{aligned} c(\alpha|\alpha) &= (c\alpha|\alpha) \\ &= (T\alpha|\alpha) \\ &= (\alpha|T\alpha) \\ &= (\alpha|c\alpha) \\ &= \bar{c}(\alpha|\alpha). \end{aligned}$$

Como $(\alpha|\alpha) \neq 0$, debemos tener que $c = \bar{c}$. Supóngase que también tenemos que $T\beta = d\beta$ con $\beta \neq 0$. Entonces

$$\begin{aligned} c(\alpha|\beta) &= (T\alpha|\beta) \\ &= (\alpha|T\beta) \\ &= (\alpha|d\beta) \\ &= \bar{d}(\alpha|\beta) \\ &= d(\alpha|\beta). \end{aligned}$$

Si $c \neq d$, entonces $(\alpha|\beta) = 0$. ∎

Debe observarse que el Teorema 15 no dice nada respecto a la existencia de valores propios o de vectores propios.

Teorema 16. *En un espacio producto interno de dimensión finita y positiva todo operador autoadjunto tiene un vector propio (no nulo).*

Demostración. Sea V un espacio producto interno de dimensión n, donde $n > 0$, y sea T un operador autoadjunto en V. Se elige una base ortonormal $\mathcal{B}$ para V y sea $A = [T]_{\mathcal{B}}$. Como $T = T^*$, tenemos que $A = A^*$. Ahora sea W el espacio de las matrices $n \times 1$ sobre C, con el producto interno $(X|Y) = Y^*X$. Entonces $U(X) = AX$ define un operador lineal autoadjunto U en W. El polinomio característico, det $(xI - A)$, es un polinomio de grado n sobre los números complejos; todo polinomio sobre C de grado positivo tiene una raíz. Así existe un número complejo c tal que det $(cI - A) = 0$. Esto quiere decir que $A - cI$ es singular, o que existe un X distinto de cero tal que $AX = cX$. Como el operador U (multiplicación por A) es autoadjunto, se sigue del Teorema 15 que c es real. Si V es un espacio vectorial real, se puede elegir X de modo

que tenga elementos reales. En efecto, entonces A y $A - cI$ tienen elementos reales y como $A - cI$ es singular, el sistema $(A - cI)X = 0$ tiene solución real X no nula. Se sigue que existe un vector α, no nulo, en V tal que $T\alpha = c\alpha$. ∎

Varios comentarios se pueden hacer respecto a la demostración.

(1) La demostración de la existencia de una X no nula tal que $AX = cX$ no tiene nada que ver con el hecho de que A fuese hermítica (autoadjunta). Ello muestra que cualquier operador lineal sobre un espacio vectorial complejo de dimensión finita tiene un vector propio. En el caso de un espacio producto interno real, el ser A autoadjunta se usa fundamentalmente para ver que cada valor propio de A es real y luego que se puede hallar una X apropiada con elementos reales.

(2) El razonamiento muestra que el polinomio característico de una matriz autoadjunta tiene coeficientes reales a pesar de que la matriz puede no tener elementos reales.

(3) La suposición de que V es de dimensión finita es necesaria para el teorema; un operador autoadjunto en un espacio producto interno de dimensión infinita puede no tener un valor propio.

Ejemplo 29. Sea V el espacio vectorial de las funciones continuas de valor complejo (o real) en el intervalo unitario, $0 \leq t \leq 1$, con el producto interno

$$(f|g) = \int_0^1 f(t)\overline{g(t)}\, dt.$$

El operador «multiplicación por t», $(Tf)(t) = tf(t)$, es autoadjunto. Supongamos que $Tf = cf$. Entonces

$$(t - c)f(t) = 0, \qquad 0 \leq t \leq 1$$

y así $f(t) = 0$ para $t \neq c$. Como f es continuo, $f = 0$. Luego T tiene valores (vectores) propios.

Teorema 17 *Sea V un espacio producto interno de dimensión finita y sea T un operador lineal sobre V. Supóngase que W es un subespacio de V que es invariante por T. Entonces el complemento ortogonal de W es invariante por T^*.*

Demostración. Recuérdese que el que W sea invariante por T no quiere decir que todo vector de W quede fijo por T; quiere decir que si α está en W entonces $T\alpha$ está en W. Sea β en $W^\perp$. Debemos mostrar que $T^*\beta$ está en $W^\perp$, esto es, que $(\alpha|T^*\beta) = 0$ para todo α en W. Si α está en W, entonces $T\alpha$ está en W, de modo que $(T\alpha|\beta) = 0$. Pero $(T\alpha|\beta) = (\alpha|T^*\beta)$. ∎

Teorema 18. *Sea V un espacio producto interno de dimensión finita y sea T un operador lineal autoadjunto sobre V. Entonces existe una base ortonormal de V en la que cada vector es un vector propio de T.*

Demostración. Se supone que $\dim V > 0$. Por el Teorema 16, T tiene un vector propio α. Sea $\alpha_1 = \alpha/\|\alpha\|$ de modo que α_1 es también un vector propio para T y $\|\alpha_1\| = 1$. Si $\dim V = 1$, se habrá concluido con la demostración.

Se procede ahora por inducción sobre la dimensión de V. Supóngase que el teorema es válido para espacios producto interno de dimensión menor que dim V. Sea W el espacio unidimensional generado por el vector α_1. La afirmación de que α_1 es un vector propio para T quiere decir simplemente que W es invariante por T. Por el Teorema 17, el complemento ortogonal $W^\perp$ es invariante por $T^* = T$. Ahora W, con el producto interno de V, es un espacio producto interno de dimensión menor en uno que la dimensión de W. Sea U el operador lineal inducido en $W^\perp$ por T, esto es, la restricción de T a $W^\perp$. Entonces U es autoadjunto y, por la hipótesis inductiva, W tiene una base ortonormal $\{\alpha_2, \ldots, \alpha_n\}$ que consiste en vectores propios de U. Ahora cada uno de estos vectores es también un vector propio de T, y como $V = W \oplus W^\perp$, concluimos que $\{\alpha_1, \ldots, \alpha_n\}$ es la deseada base para V. ∎

Corolario. *Sea A la matriz $n \times n$ hermítica (autoadjunta). Entonces existe una matriz unitaria P tal que $P^{-1}AP$ es diagonal (A es unitariamente equivalente a una matriz diagonal). Si A es una matriz simétrica real, hay una matriz ortogonal real P tal que $P^{-1}AP$ es diagonal.*

Demostración. Sea $V = C^{n \times 1}$, con el producto interno canónico, y sea T el operador lineal sobre V representado por A en la base ordenada canónica. Como $A = A^*$, tenemos que $T = T^*$. Sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ordenada ortonormal de V, tal que $T\alpha_j = c_j\alpha_j$, $j = 1, \ldots, n$. Si $D = [T]_{\mathcal{B}}$, entonces D es la matriz diagonal con elementos $c_1, \ldots, c_n$ en esa diagonal. Sea P la matriz con vectores columna $\alpha_1, \ldots, \alpha_n$. Entonces $D = P^{-1}AP$.

En el caso de que cada elemento de A sea real, se puede tomar V como R^n, con el producto interno canónico, y se repite el razonamiento. En este caso P será una matriz unitaria con elementos reales, es decir, una matriz ortogonal real. ∎

Combinando el Teorema 18 con los comentarios que hicimos al comienzo de esta sección, se tiene lo siguiente: si V es un espacio producto interno *real* de dimensión finita y T es un operador lineal sobre V, entonces V tiene una base ortonormal de vectores propios para T si, y solo si, T es autoadjunto. A esto equivale decir: si A es una matriz $n \times n$ con elementos *reales*, existe una matriz ortogonal real P tal que P^tAP es diagonal si, y solo si, $A = A^t$. No existe un resultado semejante para matrices simétricas complejas. En otras palabras, para matrices complejas hay una diferencia significativa entre las condiciones $A = A^t$ y $A = A^*$.

Habiendo visto el caso de los autoadjuntos, volvemos ahora al estudio de los operadores normales en general. Se demostrará el análogo al Teorema 18, para operadores normales, en el caso *complejo*. Hay una razón para esta restricción. Un operador normal sobre un espacio producto interno real puede no tener un vector propio no nulo. Esto es cierto, por ejemplo, para todas las rotaciones en R^2, a excepción de dos.

Teorema 19. *Sean V un espacio producto interno de dimensión finita y T un operador normal sobre V. Supóngase que α es un vector de V. Entonces α es un*

vector propio de T con valor propio c si, y solo si, α es un vector propio para T^* con valor propio $\bar{c}$.

Demostración. Supóngase que U sea un operador normal en V. Entonces $\|U\alpha\| = \|U^*\alpha\|$. En efecto, usando la condición de que $UU^* = U^*U$ se tiene que

$$\begin{aligned} \|U\alpha\|^2 = (U\alpha|U\alpha) &= (\alpha|U^*U\alpha) \\ &= (\alpha|UU^*\alpha) = (U^*\alpha|U^*\alpha) = \|U^*\alpha\|^2. \end{aligned}$$

Si c es un escalar cualquiera, el operador $U = T - cI$ es normal. En efecto, $(T - cI)^* = T^* - \bar{c}I$, con lo que es fácil de verificar que $UU^* = U^*U$. Así,

$$\|(T - cI)\alpha\| = \|(T^* - cI)\alpha\|$$

de modo que $(T - cI)\alpha = 0$ si, y solo si, $(T^* - \bar{c}I)\alpha = 0$. ▮

Definición. *Una matriz compleja $n \times n$, A, se llama* **normal** *si $AA^* = A^*A$.*

No es muy fácil entender lo que significa normal en matrices u operadores; sin embargo, con objeto de desarrollar un sentido para tal concepto, el lector podría encontrar de utilidad saber que una matriz triangular es normal si, y solo si, es diagonal.

Teorema 20. *Sean V un espacio producto interno de dimensión finita, T un operador lineal sobre V y $\mathcal{B}$ una base ortonormal de V. Supóngase que la matriz A de T en la base $\mathcal{B}$ es triangular superior. Entonces T es normal si, y solo si, A es una matriz diagonal.*

Demostración. Como $\mathcal{B}$ es una base ortogonal, A^* es la matriz de T^* en $\mathcal{B}$. Si A es diagonal, entonces $AA^* = A^*A$, y esto implica que $TT^* = T^*T$. Recíprocamente, supóngase que T sea normal y sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$. Entonces, como A es triangular superior, $T\alpha_1 = A_{11}\alpha_1$. Por el Teorema 19 esto implica que $T^*\alpha_1 = \bar{A}_{11}\alpha_1$. Por otro lado

$$\begin{aligned} T^*\alpha_1 &= \sum_j (A^*)_{j1}\alpha_j \\ &= \sum_j \bar{A}_{1j}\alpha_j. \end{aligned}$$

Por tanto, $A_{1j} = 0$, para todo $j > 1$. En particular $A_{12} = 0$, y como A es triangular superior, se sigue que

$$T\alpha_2 = A_{22}\alpha_2.$$

Así, $T^*\alpha_2 = \bar{A}_{22}\alpha_2$ y $A_{2j} = 0$ para todo $j \neq 0$. Continuando de esta manera se llega a demostrar que A es diagonal. ▮

Teorema 21. *Sea V un espacio producto interno de dimensión finita y sea T un operador lineal sobre V. Entonces existe una base ortonormal de V en la que la matriz de T es triangular superior.*

Demostración. Sea n la dimensión de V. El teorema es cierto cuando $n = 1$, y se procede por inducción en n, suponiendo que el resultado es cierto para operadores lineales en espacios producto interno complejos de dimensión $n-1$. Como V es un espacio producto interno de dimensión finita, hay un vector unitario α en V y un escalar c tales que

$$T^*\alpha = c\alpha.$$

Sea W el complemento ortogonal del subespacio generado por α y sea S la restricción de T a W. Por el Teorema 17, W es invariante por T. Así que S es un operador lineal sobre W. Como W tiene dimensión $n-1$, la hipótesis inductiva implica la existencia de una base ortonormal $\{\alpha_1, \ldots, \alpha_{n-1}\}$ de W en la cual la matriz S es triangular superior; sea $\alpha_n = \alpha$. Entonces $\{\alpha_1, \ldots, \alpha_n\}$ es una base ortonormal de V en la que la matriz de T es triangular superior. ▌

Este teorema implica el siguiente resultado para las matrices.

Corolario. *Para toda matriz compleja $n \times n$, A, existe una matriz unitaria U tal que $U^{-1}AU$ es triangular superior.*

Ahora, combinando el Teorema 21 con el Teorema 20, se tiene, sin más, el siguiente análogo al Teorema 18 para operadores normales.

Teorema 22. *Sean V un espacio producto interno complejo de dimensión finita y T un operador normal sobre V. Entonces V tiene una base ortonormal que consiste en vectores propios de T.*

También existe una interpretación matricial.

Corolario. *Para toda matriz normal A existe una matriz unitaria P tal que $P^{-1}AP$ es una matriz diagonal.*

Ejercicios

1. Para cada una de las siguientes matrices simétricas reales, A, hallar una matriz ortogonal real P tal que P^tAP sea diagonal.

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ \operatorname{sen}\theta & -\cos\theta \end{bmatrix}$$

2. ¿Es una matriz simétrica compleja autoadjunta? ¿Es normal?

3. Para

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

hay una matriz ortogonal real P tal que $P^{-1}AP = D$ sea diagonal. Hallar tal matriz diagonal D.

4. Sea $V = C^2$, con el producto interno canónico. Sea T el operador lineal sobre V representado en la base canónica por la matriz

$$A = \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}.$$

Demostrar que T es normal y hallar una base ortogonal de V que consista en vectores propios de T.

5. Dar un ejemplo de una matriz 2×2, A, tal que A^2 sea normal, pero que A no lo sea.

6. Sea T un operador normal sobre un espacio producto interno complejo de dimensión finita. Demostrar que T es autoadjunto, positivo, o unitario según que todos los valores propios de T sean reales, positivos o de valor absoluto igual a 1. (Usar el Teorema 22 para reducir el problema a uno similar respecto a matrices diagonales.)

7. Sea T un operador lineal sobre el espacio producto interno de dimensión finita V y supóngase que T es positivo y unitario. Demostrar que $T = I$.

8. Demostrar que T es normal si, y solo si, $T = T_1 + iT_2$, donde T_1 y T_2 son operadores autoadjuntos que conmutan.

9. Demostrar que una matriz simétrica real tiene una raíz cúbica simétrica; es decir, si A es simétrica real, existe una simétrica real B tal que $B^3 = A$.

10. Demostrar que toda matriz positiva es cuadrado de una matriz positiva.

11. Demostrar que un operador normal y nilpotente es el operador cero.

12. Si T es un operador normal, demostrar que los vectores propios de T, asociados a distintos valores propios, son ortogonales.

13. Sea T un operador normal sobre un espacio producto interno complejo de dimensión finita. Demostrar que existe un polinomio f, con coeficientes complejos, tal que $T^* = f(T)$. (Representar T por medio de una matriz diagonal y ver qué debe ser f.)

14. Si dos operadores normales conmutan, demostrar que su producto es normal.

9. *Operadores sobre espacios producto interno*

9.1. Introducción

La mayoría de los temas del Capítulo 8 se consideran fundamentales. Es la materia que todos deben saber. El presente capítulo es para estudiantes más avanzados o para el lector deseoso de ampliar sus conocimientos acerca de operadores sobre espacios producto interno. Con la excepción del teorema del eje principal, que esencialmente no es más que otra formulación del Teorema 18 sobre la diagonalización ortogonal de operadores autoadjuntos, y los otros resultados sobre formas de la Sección 9.2, el material que aquí se presenta es más complicado y generalmente implica mayor tecnicismo. También exigimos más del lector, como se hizo en la parte final de los Capítulos 5 y 7. Los razonamientos y demostraciones están escritos en estilo más condensado y no se dan casi ejemplos para facilitar el camino; sin embargo, hemos procurado que el lector disponga de abundantes conjuntos de ejercicios.

Las tres primeras secciones se dedican a los resultados concernientes a formas en espacios producto interno y a la relación entre formas y operadores lineales. La siguiente sección se ocupa de la teoría espectral, es decir, de las consecuencias de los Teoremas 18 y 22 del Capítulo 8 en lo que se refiere a la diagonalización de operadores autoadjuntos y normales. En la sección final proseguimos el estudio de los operadores normales tratando en particular el caso real; al hacerlo se examina lo que dice el teorema de descomposición prima del Capítulo 6 respecto a los operadores normales.

9.2. Formas sobre espacios producto interno

Si T es un operador lineal sobre un espacio producto interno de dimensión finita V, la función f definida sobre $V \times V$ por

$$f(\alpha, \beta) = (T\alpha|\beta)$$

puede considerarse como un tipo de sustituto para T. Muchas cuestiones respecto de T equivalentes a cuestiones respecto de f. En efecto, es fácil ver que f determina a T. Pues si $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ es una base ortonormal de V, entonces los elementos de la matriz de T en $\mathcal{B}$ vienen dados por

$$A_{jk} = f(\alpha_k, \alpha_j).$$

Es importante entender por qué f determina a T desde un punto de vista más abstracto. Las propiedades características de f están descritas en la siguiente definición.

Definición. *Una* **forma (sesquilineal)** *sobre un espacio vectorial real o complejo V es una función f sobre $V \times V$ con valores en el cuerpo de los escalares tal que*

(a) $$f(c\alpha + \beta, \gamma) = cf(\alpha, \gamma) + f(\beta, \gamma)$$
(b) $$f(\alpha, c\beta + \gamma) = \bar{c}f(\alpha, \beta) + f(\alpha, \gamma)$$

para todo α, β, γ de V y todos los escalares c.

Así, una forma sesquilineal es una función sobre $V \times V$ tal que $f(\alpha, \beta)$ es una función lineal de α para β fijo y una función lineal-conjugada de β para α fijo. En el caso real, $f(\alpha, \beta)$ es una función lineal de cada argumento; o sea que f es una **forma bilineal.** En el caso complejo, la forma sesquilineal f no es bilineal, a menos que $f = 0$. En el resto de este capítulo se omitirá el adjetivo «sesquilineal», salvo que sea importante tenerlo en cuenta.

Si f y g son formas sobre V y c es un escalar, es fácil probar que $cf + g$ es también una forma. De esto se concluye que toda combinación lineal de formas sobre V es a su vez una forma. Así, el conjunto de todas las formas sobre V es un subespacio del espacio vectorial de las funciones escalares sobre $V \times V$.

Teorema 1. *Sean V un espacio producto interno de dimensión finita y f una forma sobre V. Entonces existe un operador lineal único sobre V tal que*

$$f(\alpha, \beta) = (T\alpha|\beta)$$

para cualesquiera α, β de V, y la aplicación $f \rightarrow T$ es un isomorfismo del espacio de las formas sobre $L(V, V)$.

Demostración. Fijado el vector β en V, entonces $\alpha \rightarrow f(\alpha, \beta)$ es una función lineal sobre V. Por el Teorema 6 existe un vector único β' de V tal que $f(\alpha, \beta) = (\alpha|\beta')$ para todo α. Se define una función U de V en V haciendo $U\beta = \beta'$. Entonces

$$\begin{aligned} f(\alpha|c\beta + \gamma) &= (\alpha|U(c\beta + \gamma)) \\ &= \bar{c}f(\alpha, \beta) + f(\alpha, \gamma) \\ &= \bar{c}(\alpha|U\beta) + (\alpha|U\gamma) \\ &= (\alpha|cU\beta + U\gamma) \end{aligned}$$

para cualesquiera α, β, γ en V y todos los escalares c. Así, U es un operador lineal sobre V, y $T = U^*$ es un operador tal que $f(\alpha, \beta) = (T\alpha|\beta)$ para todo α y β. Si tenemos también que $f(\alpha, \beta) = (T'\alpha|\beta)$, entonces

$$(T\alpha - T'\alpha|\beta) = 0$$

para todo α y β; con lo que $T\alpha = T'\alpha$ para todo α. Así, para cada forma f existe un único operador lineal T_f tal que

$$f(\alpha, \beta) = (T_f\alpha|\beta)$$

para todo α, β de V. Si f y g son formas y c un escalar, entonces

$$\begin{aligned} (cf + g)(\alpha, \beta) &= (T_{cf+g}\alpha|\beta) \\ &= cf(\alpha, \beta) + g(\alpha, \beta) \\ &= c(T_f\alpha|\beta) + (T_g\alpha|\beta) \\ &= ((cT_f + T_g)\alpha|\beta) \end{aligned}$$

para todo α y β de V. Por tanto,

$$T_{cf+g} = cT_f + T_g$$

y $f \rightarrow T_f$ es una aplicación lineal. Para todo T en $L(V, V)$ la ecuación

$$f(\alpha, \beta) = (T\alpha|\beta)$$

define una forma tal que $T_f = T$, y $T_f = 0$ si, y solo si, $f = 0$. Así, $f \rightarrow T_f$ es un isomorfismo. ∎

Corolario. *La ecuación*

$$(f|g) = tr\ (T_f T_g^*)$$

define un producto interno sobre el espacio de las formas con la propiedad de que

$$(f|g) = \sum_{j,k} f(\alpha_k, \alpha_j)\overline{g(\alpha_k, \alpha_j)}$$

para toda base ortonormal $\{\alpha_1, \ldots, \alpha_n\}$ *de* V.

Demostración. Se sigue fácilmente del Ejemplo 3 del Capítulo 8 que $(T, U) \rightarrow$ tr (TU^*) es un producto interno sobre $L(V, V)$. Como $f \rightarrow T_f$ es un isomorfismo, el Ejemplo 6 del Capítulo 8 muestra que

$$(f|g) = \text{tr}\ (T_f T_g^*)$$

es un producto interno. Supongamos ahora que A y B son las matrices de T_f y T_g en la base ortonormal $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$. Entonces

$$A_{jk} = (T_f\alpha_k|\alpha_j) = f(\alpha_k, \alpha_j)$$

y $B_{jk} = (T_g\alpha_k|\alpha_j) = g(\alpha_k, \alpha_j)$. Como AB^* es la matriz de $T_fT_g^*$ en la base $\mathcal{B}$, se sigue que

$$(f|g) = \operatorname{tr}(AB^*) = \sum_{j,k} A_{jk}\bar{B}_{jk}. \quad \blacksquare$$

Definición. *Si f es una forma y $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ordenada arbitraria de V, la matriz A de elementos*

$$A_{jk} = f(\alpha_k, \alpha_j)$$

se llama **matriz de f en la base ordenada** $\mathcal{B}$.

Cuando $\mathcal{B}$ es una base ortonormal, la matriz de f en $\mathcal{B}$ es también la matriz de la transformación lineal T_f, pero en general éste no es el caso.

Si A es la matriz de f en la base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$, se sigue que

(9-1) $$f\left(\sum_s x_s\alpha_s, \sum_r y_r\alpha_r\right) = \sum_{r,s} \bar{y}_r A_{rs} x_s$$

para todos los escalares x_s é y_r $(1 \leq r, s \leq n)$. O sea que la matriz A tiene la propiedad de que

$$f(\alpha, \beta) = Y^*AX$$

donde X e Y son las respectivas matrices de coordenadas de α y β en la base ordenada $\mathcal{B}$.

La matriz de f en otra base

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i, \ (1 \leq j \leq n)$$

está dada por la ecuación

(9-2) $$A' = P^*AP.$$

En efecto,

$$\begin{aligned} A'_{jk} &= f(\alpha'_k, \alpha'_j) \\ &= f\left(\sum_s P_{sk}\alpha_s, \sum_r P_{rj}\alpha_r\right) \\ &= \sum_{r,s} \overline{P_{rj}} A_{rs} P_{sk} \\ &= (P^*AP)_{jk}. \end{aligned}$$

Como para las matrices unitarias es $P^* = P^{-1}$, se sigue de (9-2) que los resultados que conciernen a la equivalencia unitaria pueden aplicarse al estudio de las formas.

Teorema 2. *Sea f una forma en un espacio producto interno complejo de dimensión finita. Entonces existe una base ortonormal de V en la que la matriz de f es triangular superior.*

Demostración. Sea T el operador lineal sobre V tal que $f(\alpha, \beta) = (T\alpha|\beta)$ para todo α y β. Por el Teorema 21, existe una base ortonormal $\{\alpha_1, \ldots, \alpha_n\}$ en que la matriz de T es triangular superior. Luego,

$$f(\alpha_k, \alpha_j) = (T\alpha_k|\alpha_j) = 0$$

si $j > k$. ▮

Definición. *Una forma f sobre un espacio vectorial real, o complejo, V se llama* **hermítica** *si*

$$f(\alpha, \beta) = \overline{f(\beta, \alpha)}$$

para todo α y β en V.

Si T es un operador lineal sobre un espacio producto interno de dimensión finita V y f es la forma

$$f(\alpha, \beta) = (T\alpha|\beta)$$

entonces $\overline{f(\beta, \alpha)} = (\alpha|T\beta) = (T^*\alpha|\beta)$; así que f es hermítica si, y solo si, T es autoadjunto.

Cuando f es hermítica, $f(\alpha, \alpha)$ es real para todo α, y en los espacios complejos esta propiedad caracteriza a las formas hermíticas.

Teorema 3. *Sean V un espacio vectorial complejo y f una forma sobre V tal que f(α, α) sea real para todo α. Entonces f es hermítica.*

Demostración. Sean α y β vectores de V. Se debe demostrar que $f(\alpha, \beta) = \overline{f(\beta, \alpha)}$. En efecto,

$$f(\alpha + \beta, \alpha + \beta) = f(\alpha, \beta) + f(\alpha, \beta) + f(\beta, \alpha) + f(\beta, \beta).$$

Como $f(\alpha + \beta, \alpha + \beta)$, $f(\alpha, \alpha)$ y $f(\beta, \beta)$ son reales, el número $f(\alpha, \beta) + f(\beta, \alpha)$ es real. Por el mismo razonamiento, con $\alpha + i\beta$ en vez de $\alpha + \beta$, vemos que $-if(\alpha, \beta) + if(\beta, \alpha)$ es real. Habiendo concluido que esos dos números son reales, se les hace iguales a sus complejos conjugados y obtenemos

$$f(\alpha, \beta) + f(\beta, \alpha) = \overline{f(\alpha, \beta)} + \overline{f(\beta, \alpha)}$$
$$-if(\alpha, \beta) + if(\beta, \alpha) = i\overline{f(\alpha, \beta)} - i\overline{f(\beta, \alpha)}$$

Si multiplicamos la segunda ecuación por i y se suma a la primera ecuación, tenemos

$$2f(\alpha, \beta) = 2\overline{f(\beta, \alpha)}.$$ ▮

Corolario. *Sea T un operador lineal sobre un espacio producto interno complejo de dimensión finita. Entonces T es autoadjunto si, y solo si, (Tα|α) es real para todo α de V.*

Teorema 4 (teorema del eje principal). *Para toda forma hermítica f en un espacio producto interno de dimensión finita V, existe una base ortonormal de V en la que f está representada por una matriz diagonal con elementos reales.*

Demostración. Sea T el operador lineal tal que $f(\alpha, \beta) = (T\alpha|\beta)$ para todo α y β de V. Entonces, como $f(\alpha, \beta) = \overline{f(\beta, \alpha)}$ y $\overline{(T\beta, \alpha)} = (\alpha|T\beta)$, se sigue que

$$(T\alpha|\beta) = \overline{f(\beta, \alpha)} = (\alpha|T\beta)$$

para todo α y β; luego $T = T^*$. Por el Teorema 18 del Capítulo 8, existe una base ortonormal de V formada por los vectores característicos para T. Supóngase que $\{\alpha_1, \ldots, \alpha_n\}$ es una base ortonormal y que

$$T\alpha_j = c_j\alpha_j$$

para $1 \leq j \leq n$. Entonces

$$f(\alpha_k, \alpha_j) = (T\alpha_k|\alpha_j) = \delta_{kj}c_k$$

y por el Teorema 15 del Capítulo 8, todo c_k es real. ▮

Corolario. *En las condiciones anteriores*

$$f(\sum_j x_j\alpha_j, \sum_k y_k\alpha_k) = \sum_j c_j x_j \bar{y}_j.$$

Ejercicios

1. ¿Cuáles de las siguientes funciones f definidas para los vectores $\alpha = (x_1, x_2)$ y $\beta = (y_1, y_2)$ de C^2 son formas (sesquilineales) sobre C^2?

(a) $f(\alpha, \beta) = 1$.
(b) $f(\alpha, \beta) = (x_1 - \bar{y}_1)^2 + x_2\bar{y}_2$.
(c) $f(\alpha, \beta) = (x_1 + \bar{y}_1)^2 - (x_1 - \bar{y}_1)^2$.
(d) $f(\alpha, \beta) = x_1\bar{y}_2 - \bar{x}_2 y_1$.

2. Sea f la forma sobre R^2 definida por

$$f((x_1, y_1), (x_2, y_2)) = x_1y_1 + x_2y_2.$$

Hallar la matriz de f en cada una de las siguientes bases:

$$\{(1, 0), (0, 1)\}, \{(1, -1), (1, 1)\}, \{(1, 2), (3, 4)\}.$$

3. Sea

$$A = \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}$$

y sea g la forma (en el espacio de las matrices complejas 2×1) definida por $g(X, Y) = Y^*AX$. ¿Es g un producto interior?

4. Sea V un espacio vectorial complejo y sea f una forma (sesquilineal) simétrica sobre V: $f(\alpha, \beta) = f(\beta, \alpha)$. ¿Qué es f?

5. Sea f la forma sobre R^2 dada por

$$f((x_1, x_2), (y_1, y_2)) = x_1y_1 + 4x_2y_2 + 2x_1y_2 + 2x_2y_1.$$

Hallar una base ordenada en la que f esté representada por una matriz diagonal.

6. La forma f se llama **no degenerada** (a la izquierda) si 0 es el único vector α tal que $f(\alpha, \beta) = 0$ para todo β. Sea f una forma sobre un espacio producto interno V. Demostrar que f es no degenerada si, y solo si, el operador lineal asociado T_f (Teorema 1) es no singular.

7. Sea f una forma en un espacio vectorial de dimensión finita V. Se hace referencia a la definición de la no degeneración a la izquierda dada en el Ejercicio 6. Definir la no de-

generación a la derecha y demostrar que la forma f es no degenerada a la izquierda si, y solo si, es no degenerada a la derecha.

8. Sea f una forma no degenerada (Ejercicios 6 y 7) en un espacio V de dimensión finita. Sea L un funcional lineal sobre V. Demostrar que existe un vector β, y solo uno, en V tal que $L(\alpha) = f(\alpha, \beta)$ para todo α.

9. Sea f una forma no degenerada en un espacio V de dimensión finita. Mostrar que cada operador lineal S tiene un «adjunto respecto de f», es decir, un operador S' tal que $f(S\alpha, \beta) = f(\alpha, S'\beta)$ para todo α, β.

9.3. *Formas positivas*

En esta sección se examinarán formas (sesquilineales) no negativas y sus relaciones con un producto interno dado sobre el espacio vectorial principal.

Definiciones. *Una forma f en un espacio vectorial real o complejo V es* **no negativa** *si es hermítica y $f(\alpha, \alpha) \geq 0$ para todo α en V. La forma f es* **positiva** *si f es hermítica y $f(\alpha, \alpha) > 0$ para todo $\alpha \neq 0$.*

Una forma positiva en V es simplemente un producto interno sobre V. Una forma no negativa satisface todas las propiedades de un producto interno, con la excepción de que hay vectores no nulos que pueden ser «ortogonales» consigo mismos.

Sea f una forma en un espacio V de dimensión finita. Sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ordenada de V y sea A la matriz de f en la base $\mathcal{B}$, esto es, $A_{jk} = f(\alpha_k, \alpha_j)$. Si $\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n$, entonces

$$\begin{aligned} f(\alpha, \alpha) &= f(\sum_j x_j\alpha_j, \sum_k x_k\alpha_k) \\ &= \sum_j \sum_k x_j\bar{x}_k f(\alpha_j, \alpha_k) \\ &= \sum_j \sum_k A_{kj}x_j\bar{x}_k. \end{aligned}$$

Así, pues, se ve que f es no negativa si, y solo si,

$$A = A^*$$

y

(9-3) $$\sum_j \sum_k A_{kj}x_j\bar{x}_k \geq 0 \quad \text{para todos los escalares } x_1, \ldots, x_n.$$

Para que f sea positiva, la desigualdad (9-3) debe ser estricta para todo $(x_1, \ldots, x_n) \neq 0$. Las condiciones deducidas dicen que f es una forma positiva en V si, y solo si, la función

$$g(X, Y) = Y^*AX$$

es una forma positiva en el espacio de las matrices columnas sobre el cuerpo escalar.

Teorema 5. *Sea F el cuerpo de los números reales o el cuerpo de los números complejos. Sea A una matriz $n \times n$ sobre F. La función g definida por*

$$g(X, Y) = Y^*AX \tag{9-4}$$

es una forma positiva en el espacio $F^{n \times 1}$ si, y solo si, existe una matriz inversible P, con elementos en F, tal que $A = P^*P$.

Demostración. Para cualquier matriz $n \times n$, A, la función g de (9-4) es una forma en el espacio de las matrices columnas. Se demostrará que g es positiva si, y solo si, $A = P^*P$. Primeramente, supóngase que $A = P^*P$. Entonces g es hermítica y

$$\begin{aligned} g(X, X) &= X^*P^*PX \\ &= (PX)^*PX \\ &\geq 0. \end{aligned}$$

Si P es inversible y $X \neq 0$, entonces $(PX)^*PX > 0$. Ahora supóngase que g es una forma positiva en el espacio de las matrices columnas. Entonces es un producto interno y, por tanto, existen matrices columnas $Q_1, \ldots, Q_n$ tales que

$$\begin{aligned} \delta_{jk} &= g(Q_j, Q_k) \\ &= Q_k^*AQ_j. \end{aligned}$$

Pero esto dice, justamente, que si Q es la matriz con columnas $Q_1, \ldots, Q_n$, entonces $Q^*AQ = I$. Como $\{Q_1, \ldots, Q_n\}$ es una base, Q es inversible. Con $P = Q^{-1}$ se tiene $A = P^*P$. ▌

En la práctica no es tan fácil verificar si una matriz dada, A, satisface el criterio de positividad que se ha dado hasta ahora. Una consecuencia del último teorema es que si g es positiva, entonces $\det A > 0$, ya que $\det A = \det (P^*P) = \det P^* \det P = |\det P|^2$. El que $\det A > 0$ no quiere decir que sea suficiente para garantizar que g es positiva; sin embargo, hay n determinantes asociados con A que tienen esta propiedad: si $A = A^*$ y si cada uno de estos determinantes es positivo, entonces g es una forma positiva.

Definición. *Sea A una matriz $n \times n$ sobre el cuerpo F. Los* **menores principales** *de A son los escalares $\Delta_k(A)$ definidos por*

$$\Delta_k(A) = det \begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & & \vdots \\ A_{k1} & \cdots & A_{kk} \end{bmatrix}, \qquad 1 \leq k \leq n.$$

Lema. *Sea A una matriz inversible $n \times n$ con elementos en un cuerpo F. Las dos afirmaciones siguientes son equivalentes.*

(a) *Existe una matriz triangular superior P con $P_{kk} = 1$ $(1 \leq k \leq n)$ tal que la matriz $B = AP$ es triangular inferior.*

(b) *Los menores principales de A son todos distintos de* 0.

Demostración. Sea P cualquier matriz $n \times n$ y hágase $B = AP$. Entonces

$$B_{jk} = \sum_r A_{jr}P_{rk}.$$

Si P es triangular superior y $P_{kk} = 1$ para todo k, entonces

$$\sum_{r=1}^{k-1} A_{jr}P_{rk} = B_{jk} - A_{kk}, \qquad k > 1.$$

Ahora bien, B es triangular inferior siempre que $B_{jk} = 0$ para $j < k$. Así, pues, B será triangular inferior si, y solo si,

$$\text{(9-5)} \qquad \sum_{r=1}^{k-1} A_{jr}P_{rk} = -A_{kk}, \qquad \begin{array}{l} 1 \leq j \leq k-1 \\ 2 \leq k \leq n. \end{array}$$

Con lo que vemos que la afirmación (a) del lema es equivalente a la afirmación de que existen escalares P_{rk}, $1 \leq r \leq k$, $1 \leq k \leq n$, que satisfacen (9-5) y $P_{kk} = 1$, $1 \leq k \leq n$.

En (9-5), para todo $k > 1$ tenemos un sistema de $k - 1$ ecuaciones lineales para las incógnitas $P_{1k}, P_{2k}, \ldots, P_{k-1,k}$. La matriz de coeficientes de ese sistema es

$$\begin{bmatrix} A_{11} & \cdots & A_{1,k-1} \\ \vdots & & \vdots \\ A_{k-1} & \cdots & A_{k-1,k-1} \end{bmatrix}$$

y su determinante es el menor principal $\Delta_{k-1}(A)$. Si cada $\Delta_{k-1}(A) \neq 0$, el sistema (9-5) tiene soluciones únicas. Hemos demostrado que la afirmación (b) implica la afirmación (a) y que la matriz P es única.

Ahora supóngase que se verifica (a). Entonces, como se verá,

$$\text{(9-6)} \qquad \begin{aligned} \Delta_k(A) &= \Delta_k(B) \\ &= B_{11}B_{22}\cdots B_{kk}, \qquad k = 1, \ldots, n. \end{aligned}$$

Para verificar (9-6), sean $A_1, \ldots, A_n$ y $B_1, \ldots, B_n$ las columnas de A y B, respectivamente. Entonces

$$\text{(9-7)} \qquad \begin{aligned} B_1 &= A_1 \\ B_r &= \sum_{j=1}^{r-1} P_{jr}A_j + A_r, \qquad r > 1. \end{aligned}$$

Se deja fijo k, $1 \leq k \leq n$. Por (9-7) se ve que la r-ésima columna de la matriz

$$\begin{bmatrix} B_{11} & \cdots & B_{kk} \\ \vdots & & \vdots \\ B_{k1} & \cdots & B_{kk} \end{bmatrix}$$

se obtiene sumando a la r-ésima columna de

$$\begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & & \vdots \\ A_{k1} & \cdots & A_{kk} \end{bmatrix}$$

una combinación lineal de sus otras columnas. Tales operaciones no alteran los determinantes. Lo que demuestra (9-6), a excepción de la observación trivial de que por ser B triangular, $\Delta_k(B) = B_{11}, \ldots, B_{kk}$. Como A y P son inversibles, B es inversible. Por tanto,

$$\Delta(B) = B_{11} \cdots B_{nn} \neq 0$$

y así $\Delta_k(A) \neq 0$, $k = 1, \ldots, n$. ▮

Teorema 6. *Sea f una forma en un espacio vectorial V de dimensión finita y sea A la matriz de f en una base ordenada $\mathcal{B}$. Entonces f es una forma positiva si, y solo si, $A = A^*$ y los menores principales de A son todos positivos.*

Demostración. Hagamos la mitad interesante del teorema. Supóngase que $A = A^*$ y que $\Delta_k(A) > 0$, $1 \leq k \leq n$. Por el lema existe una matriz triangular superior (única) P con $P_{kk} = 1$ tal que $B = AP$ es triangular inferior. La matriz P^* es triangular inferior, con lo que $P^*B = P^*AP$ es también triangular inferior. Como A es autoadjunta, la matriz $D = P^*AP$ es autoadjunta. Por el mismo razonamiento que llevó a (9-6),

$$\begin{aligned}\Delta_k(D) &= \Delta_k(P^*B) \\ &= \Delta_k(B) \\ &= \Delta_k(A).\end{aligned}$$

Como D es diagonal, sus menores principales son

$$\Delta_k(D) = D_{11} \cdots D_{kk}.$$

De $\Delta_k(D) > 0$, $1 \leq k \leq n$, se obtiene que $D_{kk} > 0$ para todo k.

Si A es la matriz de la forma f en la base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$, entonces $D = P^*AP$ es la matriz de f en la base $\{\alpha'_1, \ldots, \alpha'_n\}$ definida por

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Véase (9-2). Como D es diagonal con elementos positivos en su diagonal, es obvio que

$$X^*DX > 0, \qquad X \neq 0$$

de lo que se sigue que f es una forma positiva.

Ahora supóngase que f es una forma positiva. Se sabe que $A = A^*$. ¿Cómo se demuestra que $\Delta_k(A) > 0$, $1 \leq k \leq n$? Sea V_k el subespacio generado por $\alpha_1, \ldots, \alpha_k$ y sea f_k la restricción de f a $V_k \times V_k$. Evidentemente f_k es una forma positiva en V_k y, en la base $\{\alpha_1, \ldots, \alpha_k\}$, está representada por la matriz

$$\begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & & \vdots \\ A_{k1} & \cdots & A_{kk} \end{bmatrix}.$$

Como consecuencia del Teorema 5 se observó que la positividad de la forma implica que el determinante de cualquier matriz que la represente es positiva. ▮

Quedan algunos comentarios por hacer para completar el estudio de la relación entre formas positivas y matrices. ¿Qué es lo que caracteriza las matrices que representan a las formas positivas? Si f es una forma en un espacio vectorial complejo y A es la matriz de f en cierta base ordenada, entonces f será positiva si, y solo si, $A = A^*$ y

(9-8) $X^*AX > 0$, para todo $X \neq 0$ complejo

Se sigue del Teorema 3 que la condición $A = A^*$ es redundante, es decir, que (9-8) implica $A = A^*$. Por otro lado, si se tiene un espacio vectorial real, la forma f será positiva si, y solo si, $A = A^*$ y

(9-9) $X^tAX > 0$, para todo $X \neq 0$ real.

Hay que recalcar que si una matriz real A satisface (9-9) no se desprende de eso que $A = A^t$. Lo que sí es cierto es que si se cumplen $A = A^t$ y (9-9), entonces también se verifica (9-8). Y es porque

$$\begin{aligned}(X + iY)^*A(X + iY) &= (X^t - iY^t)A(X + iY)\\ &= X^tAX + Y^tAY + i[X^tAY - Y^tAX]\end{aligned}$$

y si $A = A^t$, entonces $Y^tAX = X^tAY$.

Si A es una matriz $n \times n$ de elementos complejos y si A satisface (9-9), se dirá que A es una **matriz positiva.** Los comentarios que acabamos de hacer pueden resumirse diciendo: en los casos real o complejo, una forma f es positiva si, y solo si, su matriz en cierta base (de hecho, en toda) ordenada es una matriz positiva.

Supóngase ahora que V es un espacio producto interno de dimensión finita. Sea f una forma no negativa en V. Existe un único operador autoadjunto T sobre V tal que

(9-10) $f(\alpha, \beta) = (T\alpha|\beta).$

y T tiene, además, la propiedad de que $(T\alpha|\alpha) \geq 0$.

Definición. *Un operador lineal T sobre un espacio producto interno de dimensión finita V es* **no negativo** *si $T = T^*$ y $(T\alpha|\alpha) \geq 0$ para todo α de V. Un operador lineal es* **positivo** *si $T = T^*$ y $(T\alpha|\alpha) > 0$ para todo $\alpha \neq 0$.*

Si V es un espacio vectorial (real o complejo) de dimensión finita, y si $(\cdot|\cdot)$ es un producto interno sobre V, hay una clase asociada de operadores lineales positivos sobre V. Mediante (9-10) existe una correspondencia biunívoca entre esta clase de operadores positivos y la colección de todas las formas positivas en V. Se usarán los ejercicios de esta sección para destacar la relación entre operadores positivos, formas positivas y matrices positivas. El siguiente resumen puede ser de utilidad.

Si A es una matriz $n \times n$ sobre el cuerpo de los números complejos, las siguientes afirmaciones son equivalentes.

(1) A es positiva, es decir, $\sum_j \sum_k A_{kj} x_j \bar{x}_k > 0$ siempre que $x_1, \ldots, x_n$ sean números complejos, no todos 0.

(2) $(X|Y) = Y^*AY$ es un producto interno en el espacio de las matrices complejas $n \times 1$.

(3) Respecto al producto interno canónico $(X|Y) = Y^*X$ de matrices $n \times 1$, el operador lineal $X \rightarrow AX$ es positivo.

(4) $A = P^*P$ para alguna matriz inversible $n \times n$, P, sobre C.

(5) $A = A^*$, y los menores principales de A son positivos.

Si todo elemento de A es real, éstas equivalen a las siguientes:

(6) $A = A^*$, y $\sum_j \sum_k A_{kj}x_jx_k > 0$ siempre que $x_1, \ldots, x_n$ sean números reales no todos nulos.

(7) $(X|Y) = Y^tAX$ es un producto interno sobre el espacio de las matrices $n \times 1$ reales.

(8) Respecto al producto interno canónico $(X|Y) = Y^tX$, sobre las matrices reales $n \times 1$, el operador lineal $X \rightarrow AX$ es positivo.

(9) Existe una matriz inversible $n \times n$, P, con elementos reales tal que $A = P^tP$.

Ejercicios

1. Sea $V = C^2$, con el producto interno canónico. ¿Para cuáles vectores α de V hay un operador lineal positivo T tal que $\alpha = T\epsilon_1$?

2. Sea $V = R^2$, con el producto interno canónico. Si θ es un número real, sea T el operador lineal «rotación θ»,

$$T_\theta(x_1, x_2) = (x_1 \cos \theta - x_2 \operatorname{sen} \theta, x_1 \operatorname{sen} \theta + x_2 \cos \theta).$$

¿Para qué valores de θ es T_θ un operador positivo?

3. Sea V el espacio de las matrices $n \times 1$ sobre C, con el producto interno $(X|Y) = Y^*GX$ (donde G es una matriz $n \times n$ tal que éste sea un producto interno). Sean A una matriz $n \times n$ y T un operador lineal $T(X) = AX$. Hallar T^*. Si Y es un elemento fijo de Y, hallar el elemento Z de V que determina el funcional lineal $X \rightarrow Y^*X$. En otras palabras, hallar Z tal que $Y^*X = (X|Z)$ para todo X de V.

4. Sea V un espacio producto interno de dimensión finita. Si T y U son operadores lineales positivos sobre V, demostrar que $(T + U)$ es positivo. Dar un ejemplo que muestre que TU no tiene que ser positivo necesariamente.

5. Sea
$$A = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}.$$

(a) Demostrar que A es positiva.

(b) Sea V el espacio de las matrices reales 2×1, con el producto interno $(X|Y) = Y^tAX$. Hallar una base ortonormal de V aplicando el proceso de Gram-Schmidt a la base $\{X_1, X_2\}$ definida por

$$X_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad X_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

(c) Hallar una matriz real inversible 2×1, P, tal que $A = P^tP$.

6. ¿Cuáles de las siguientes matrices son positivas?

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1+i \\ 1-i & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 & 1 \\ 2 & -1 & 1 \\ 3 & -1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}.$$

7. Dar un ejemplo de una matriz $n \times n$ que tenga todos sus menores principales positivos, pero que no sea una matriz positiva.

8. ¿Define $((x_1, x_2)|(y_1, y_2)) = x\bar{y}_1 + 2x_2\bar{y}_1 + 2x_1\bar{y}_2 + x_2\bar{y}_2$ un producto interno sobre C^2?

9. Demostrar que todo elemento de la diagonal principal de una matriz positiva es positivo.

10. Sea V un espacio producto interno de dimensión finita. Si T y U son operadores lineales sobre V, se escribe $T < U$ si $U - T$ es un operador positivo. Demostrar lo siguiente:
(a) $T < U$ y $U < T$ es imposible.
(b) Si $T < U$ y $U < S$, entonces $T < S$.
(c) Si $T < U$ y $0 < S$, no es necesariamente $ST < SU$.

11. Sean V un espacio producto interno de dimensión finita y E la proyección ortogonal de V sobre algún subespacio.
(a) Demostrar que, para cualquier número positivo c, el operador $cI + E$ es positivo.
(b) Expresar por E un operador lineal autoadjunto T tal que $T^2 = I + E$.

12. Sea n un entero positivo y A la matriz $n \times n$

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n-1} \end{bmatrix}.$$

Demostrar que A es positiva.

13. Sea A una matriz $n \times n$ autoadjunta. Demostrar que existe un número real c tal que la matriz $cI + A$ es positiva.

14. Demostrar que el producto de los operadores lineales positivos es positivo si, y solo si, conmutan.

15. Sean S y T operadores positivos. Demostrar que todo valor propio de ST es positivo.

9.4. Más sobre formas

Esta sección contiene dos resultados que dan información más detallada acerca de formas (sesquilineales).

Teorema 7. *Sean f una forma en un espacio vectorial real, o complejo, V y $\{\alpha_1, \ldots, \alpha_r\}$ una base del subespacio W de V de dimensión finita. Sea M la matriz $r \times r$ de elementos*

$$M_{jk} = f(\alpha_k, \alpha_j)$$

y W' el conjunto de todos los vectores β de V tales que $f(\alpha, \beta) = 0$ para todo α de W. Entonces W' es un subespacio de V y $W \cap W' = \{0\}$ si, y solo si, MA es inversible. Si este es el caso, $V = W + W'$

Demostración. Si β y γ son vectores de W' y c es un escalar, entonces para todo α en W

$$\begin{aligned} f(\alpha, c\beta + \gamma) &= \bar{c}f(\alpha, \beta) + f(\alpha, \gamma) \\ &= 0. \end{aligned}$$

Luego, W' es un subespacio de V.

Supóngase ahora que $\alpha = \sum_{k=1}^{r} x_k\alpha_k$ y que $\beta = \sum_{j=1}^{r} y_j\alpha_j$. Entonces

$$\begin{aligned} f(\alpha, \beta) &= \sum_{j,k} \bar{y}_j M_{jk} x_k \\ &= \sum_k \Big(\sum_j \bar{y}_j M_{jk}\Big) x_k. \end{aligned}$$

Se sigue de esto que $W \cap W' \neq \{0\}$ si, y solo si, el sistema homogéneo

$$\sum_{j=1}^{r} \bar{y}_j M_{jk} = 0, \qquad 1 \leq k \leq r$$

tiene una solución no trivial $(y_1, \ldots, y_r)$. Luego $W \cap W' = \{0\}$ si, y solo si, M^* es inversible. Pero ser M^* inversible es equivalente a ser M inversible.

Supóngase que M es inversible y sea

$$A = (M^*)^{-1} = (M^{-1})^*.$$

Defínase g_i sobre V por la ecuación

$$g_j(\beta) = \sum_{k=1}^{r} A_{jk}\overline{f(\alpha_k, \beta)}.$$

Entonces

$$\begin{aligned} g_j(c\beta + \gamma) &= \sum_k A_{jk}\overline{f(\alpha_k, c\beta + \gamma)} \\ &= c\sum_k A_{jk}f(\alpha_k, \beta) + \sum_k A_{jk}f(\alpha_k, \gamma) \\ &= cg_j(\beta) + g_j(\gamma). \end{aligned}$$

Luego toda g_i es una función lineal sobre V. Así, pues, se puede definir un operador lineal E sobre V haciendo

$$E\beta = \sum_{j=1}^{r} g_j(\beta)\alpha_j.$$

Como

$$\begin{aligned} g_j(\alpha_n) &= \sum_k A_{jk}\overline{f(\alpha_k, \alpha_n)} \\ &= \sum_k A_{jk}(M^*)_{kn} \\ &= \delta_{jn} \end{aligned}$$

se sigue que $E(\alpha_n) = \alpha_n$ para $1 \leq n \leq r$. Esto implica que $E\alpha = \alpha$ para todo α de W. Por tanto, E aplica V sobre W y $E^2 = E$. Si β es un vector arbitrario de V, entonces

$$\begin{aligned} f(\alpha_n, E\beta) &= f\left(\alpha_n, \sum_j g_j(\beta)\alpha_j\right) \\ &= \sum_j \overline{g_j(\beta)} f(\alpha_n, \alpha_j) \\ &= \sum_j \left(\sum_k \overline{A}_{jk} f(\alpha_k, \beta)\right) f(\alpha_n, \alpha_j). \end{aligned}$$

Como $A^* = M^{-1}$, se sigue que

$$\begin{aligned} f(\alpha_n, E\beta) &= \sum_k \left(\sum_j (M^{-1})_{kj} M_{jn}\right) f(\alpha_k, \beta) \\ &= \sum_k \delta_{kn} f(\alpha_k, \beta) \\ &= f(\alpha_n, \beta). \end{aligned}$$

Esto implica $f(\alpha, E\beta) = f(\alpha, \beta)$ para todo α de W. Luego,

$$f(\alpha, \beta - E\beta) = 0$$

para todo α de W y todo β de V. Así, $I - E$ aplica V en W'. La igualdad

$$\beta = E\beta + (I - E)\beta$$

muestra que $V = W + W'$. Es de mencionar un último punto. Como $W \cap W' = \{0\}$, todo vector de V es suma única de un vector de W y un vector de W'. Si β está en W' se sigue que $E\beta = 0$. Luego $I - E$ aplica V sobre W'. ∎

La proyección E construida en la demostración puede caracterizarse de la siguiente manera: $E\beta = \alpha$ si, y solo si, α está en W y $\beta - \alpha$ pertenece a W'. Así, E es independiente de la base de W utilizada en su construcción. Luego E se puede tratar como la **proyección de** V sobre W que está **determinada por** la descomposición en suma directa

$$V = W \oplus W'.$$

Obsérvese que E es una proyección ortogonal si, y solo si, $W' = W^\perp$.

Teorema 8. *Sean f una forma en un espacio vectorial real, o complejo, V y A la matriz de f en la base ordenada $\{\alpha_1, \ldots, \alpha_n\}$ de V. Supóngase que los menores principales de A son todos distintos de 0. Entonces existe una única matriz triangular superior P con $P_{kk} = 1$ $(1 \leq k \leq n)$ tal que*

$$P^*AP$$

es triangular superior.

Demostración. Como $\Delta_k(A^*) = \overline{\Delta_k(A)}$ $(1 \leq k \leq n)$, los menores principales de A^* son todos distintos de 0. Luego, por el lema utilizado en la demostración del Teorema 6, existe una matriz triangular superior P con $P_{kk} = 1$

tal que A^*P es triangular inferior. Por tanto, $P^*A = (A^*P)^*$ es triangular superior. Como el producto de las matrices triangulares superiores es también triangular superior, se sigue que P^*AP es triangular superior. Esto muestra la existencia, pero no la unicidad de P. Sin embargo, hay otro razonamiento, más geométrico, que puede usarse para demostrar la existencia y unicidad de P.

Sean W_k el subespacio generado por $\alpha_1, \ldots, \alpha_k$ y W'_k el conjunto de todos los β de V tales que $f(\alpha, \beta) = 0$ para todo α de W_k. Como $\Delta_k(A) \neq 0$, la matriz $k \times k$, M, de elementos

$$M_{ij} = f(\alpha_j, \alpha_i) = A_{ij}$$

$(1 \leq i, j \leq k)$ es inversible. Por el Teorema 7

$$V = W_k \oplus W'_k.$$

Sea E_k la proyección de V sobre W_k determinada por esta descomposición, y sea $E_0 = 0$. Sea

$$\beta_k = \alpha_k - E_{k-1}\alpha_k, \qquad (1 \leq k \leq n).$$

Entonces $\beta_1 = \alpha_1$ y $E_{k-1}\alpha_k$ pertenece a W_{k-1} para $k > 1$. Así, cuando $k > 1$, existen escalares únicos P_{jk} tales que

$$E_{k-1}\alpha_k = -\sum_{j=1}^{k-1} P_{jk}\alpha_j.$$

Haciendo $P_{kk} = 1$ y $P_{jk} = 0$ para $j > k$, se tiene entonces una matriz $n \times n$ triangular superior P con $P_{kk} = 1$ y

$$\beta_k = \sum_{j=1}^{k} P_{jk}\alpha_j$$

para $k = 1, \ldots, n$. Supóngase que $1 \leq i \leq k$. Entonces β_i está en W_i y $W_i \subset W_{k-1}$. Como β_k pertenece a W'_{k-1}, se sigue que $f(\beta_i, \beta_k) = 0$. Sea B la matriz de f en la base ordenada $\{\beta_1, \ldots, \beta_n\}$. Entonces

$$B_{ki} = f(\beta_i, \beta_k)$$

con lo que $B_{ki} = 0$ cuando $k > i$. Así B es triangular superior. Por otro lado,

$$B = P^*AP.$$

Recíprocamente, supóngase que P es una matriz triangular superior con $P_{kk} = 1$ tal que P^*AP es triangular superior. Se hace

$$\beta_k = \sum_j P_{jk}\alpha_j, \qquad (1 \leq k \leq n).$$

Entonces $\{\beta_1, \ldots, \beta_k\}$ es evidentemente una base de W_k. Supóngase que $k > 1$. Entonces $\{\beta_1, \ldots, \beta_{k-1}\}$ es una base para W_{k-1}, y como $f(\beta_i, \beta_k) = 0$ cuando $i < k$, se ve que β_k es un vector de W'_{k-1}. La ecuación que define β_k implica

$$\alpha_k = -\left(\sum_{j=1}^{k-1} P_{jk}\alpha_j\right) + \beta_k.$$

Ahora bien, $\sum_{j=k}^{k-1} P_{jk}\alpha_j$ pertenece a W_{k-1} y β_k está en W'_{k-1}. Por tanto, $P_{1k}, \ldots, P_{k-1k}$ son los únicos escalares tales que

$$E_{k-1}\alpha_k = -\sum_{j=1}^{k-1} P_{jk}\alpha_j$$

con lo que P es la matriz construida anteriormente. ∎

9.5. Teoría espectral

En esta sección prosiguen las consecuencias de los Teoremas 18 y 22 del Capítulo 8 referentes a la diagonalización de operadores autoadjuntos y normales.

Teorema 9 (teorema espectral). *Sea T un operador normal sobre un espacio producto interno complejo de dimensión finita V o un operador autoadjunto sobre un espacio producto interno real de dimensión finita V. Sean $c_1, \ldots, c_k$ los valores propios distintos de T. Sea W_j el espacio propio asociado a c_j y E_j la proyección ortogonal de W sobre W_j. Entonces W_j es ortogonal a W_i si $i \neq j$, V es la suma directa de $W_1, \ldots, W_k$ y*

(9-11) $$T = c_1E_1 + \cdots + c_kE_k.$$

Demostración. Sea α un vector de W_j, β un vector de W_i y supóngase que $i \neq j$. Entonces $c_j(\alpha|\beta) = (T\alpha|\beta) = (\alpha|T^*\beta) = (\alpha|\bar{c}_i\beta)$. Luego $(c_j - c_i)(\alpha|\beta) = 0$ y, como $c_j - c_i \neq 0$, se sigue que $(\alpha, \beta) = 0$. Así W_j es ortogonal a W_i si $i \neq j$. Del hecho de que V tiene una base ortonormal de vectores propios (véanse Teoremas 18 y 22, del Capítulo 8), se sigue que $V = W_1 + \cdots + W_k$. Si α_j pertenece a V_j $(1 \leq j \leq k)$ y $\alpha_1 + \cdots + \alpha_k = 0$, entonces

$$\begin{aligned} 0 &= (\alpha_i|\sum_j \alpha_j) = \sum_j (\alpha_i|\alpha_j) \\ &= \|\alpha_i\|^2 \end{aligned}$$

para todo i, con lo que V es la suma directa de $W_1, \ldots, W_k$. Por tanto, $E_1 + \cdots + E_k = I$ y

$$\begin{aligned} T &= TE_1 + \cdots + TE_k \\ &= c_1E_1 + \cdots + c_kE_k. \end{aligned}$$ ∎

La descomposición (9-11) se llama **descomposición espectral** *de T*. Esta terminología proviene en parte de aplicaciones físicas que han hecho que se defina el **espectro** de un operador lineal sobre un espacio vectorial de dimensión finita como el conjunto de valores propios para el operador. Es importante observar que las proyecciones ortogonales $E_1, \ldots, E_k$ están asociadas canónicamente con T; en realidad, son polinomios en T.

Corolario. *Si $e_j = \prod_{i \neq j} \left(\frac{x - c_i}{c_j - c_i}\right)$, entonces $E_j = e_j(T)$ para $1 \leq j \leq k$.*

Demostración. Como $E_iE_j = 0$ si $i \neq j$, se sigue que

$$T^2 = c_1^2E_1 + \cdots + c_k^2E_k$$

y por un sencillo razonamiento inductivo que

$$T^n = c_1^nE_1 + \cdots + c_k^nE_k$$

para todo entero $n \geq 0$. Para un polinomio arbitrario

$$f = \sum_{n=0}^{r} a_nx^n$$

se tiene

$$\begin{aligned} f(T) &= \sum_{n=0}^{r} a_nT^n \\ &= \sum_{n=0}^{r} a_n \sum_{j=1}^{k} c_j^nE_j \\ &= \sum_{j=1}^{k} \left(\sum_{n=0}^{r} a_nc_j^n\right) E_j \\ &= \sum_{j=1}^{k} f(c_j)E_j. \end{aligned}$$

Como $e_j(c_m) = \delta_{jm}$, se sigue que $e_j(T) = E_j$. ∎

Por estar $E_1, \ldots, E_k$ asociados canónicamente con T y

$$I = E_1 + \cdots + E_k$$

La familia de proyecciones $\{E_1, \ldots, E_k\}$ se llama **descomposición de la identidad definida por** T.

Cabe hacer un comentario respecto a la demostración del teorema espectral. Se derivó el teorema usando los Teoremas 18 y 22 del Capítulo 8 sobre la diagonalización de operadores autoadjuntos y normales. Hay otra demostración más algebraica, en la que debe demostrarse primero que el polinomio minimal de un operador normal es un producto de factores primos distintos. Luego se procede como en la demostración del teorema de la descomposición prima (Teorema 12, Capítulo 6). Daremos una demostración así en la siguiente sección.

En diversas aplicaciones es necesario saber si se pueden calcular ciertas funciones de operadores o matrices, v.gr., raíces cuadradas. Esto se puede hacer muy sencillamente para operadores normales diagonalizables.

Definición. *Sea T un operador normal diagonalizable sobre un espacio producto interno de dimensión finita V y sea*

$$T = \sum_{j=1}^{k} c_jE_j$$

su resolución espectral. Supóngase que f es una función cuyo dominio incluye

el espectro de T, que tenga valores en el cuerpo de los escalares. Entonces el operador lineal $f(T)$ está definido por la igualdad

(9-12) $$f(T) = \sum_{j=1}^{k} f(c_j)E_j.$$

Teorema 10. *Sea T un operador normal diagonalizable con espectro S en un espacio producto interno de dimensión finita V. Supóngase que f es una función cuyo dominio contiene a S y que toma valores en el cuerpo de los escalares. Entonces $f(T)$ es un operador normal diagonalizable de espectro $f(S)$. Si U es una aplicación unitaria de V sobre V' y $T' = UTU^{-1}$, entonces S es el espectro de T' y*

$$f(T') = Uf(T)U^{-1}$$

Demostración. La normalidad de $f(T)$ se desprende de un sencillo cálculo a partir de (9-12) y de que

$$f(T)^* = \sum_j \overline{f(c_j)}E_j.$$

Además, es evidente que para todo α en $E_j(V)$

$$f(T)\alpha = f(c_j)\alpha.$$

Así, el conjunto $f(S)$ de todos los $f(c)$ con c en S está contenido en el espectro de $f(T)$. Recíprocamente, supóngase que $\alpha \neq 0$ y que

$$f(T)\alpha = b\alpha.$$

Entonces $\alpha = \sum_j E_j\alpha$ y

$$\begin{aligned} f(T)\alpha &= \sum_j f(T)E_j\alpha \\ &= \sum_j f(c_j)E_j\alpha \\ &= \sum_j bE_j\alpha. \end{aligned}$$

Luego

$$\begin{aligned} \|\sum_j (f(c_j) - b)E_j\alpha\|^2 &= \sum_j |f(c_j) - b|^2\|E_j\alpha\|^2 \\ &= 0. \end{aligned}$$

Por tanto, $f(c_j) = b$ o $E_j\alpha = 0$. Por hipótesis, $\alpha \neq 0$, así que existe un índice i para el que $E_i\alpha \neq 0$. Se sigue que $f(c_i) = b$ y por consiguiente que $f(S)$ es el espectro de $f(T)$. Supóngase, en efecto, que

$$f(S) = \{b_1, \ldots, b_r\}$$

donde $b_m \neq b_n$ si $m \neq n$. Sea X_m el conjunto de índices i tales que $1 \leq i \leq k$ y $f(c_i) = b_m$. Sea $P_m = \sum_i E_i$ la suma extendida a los índices i de X_m. Entonces P_m es la proyección ortogonal de V sobre el espacio de los vectores propios que corresponden a los valores propios b_m de $f(T)$, y

$$f(T) = \sum_{m=1}^{r} b_mP_m$$

es la descomposición espectral de $f(T)$.

Supóngase ahora que U es una transformación unitaria de V sobre V' y que $T' = UTU^{-1}$. Entonces la igualdad

$$T\alpha = c\alpha$$

vale si, y solo si,

$$T'U\alpha = cU\alpha.$$

Así que S es el espectro de T', y U aplica todo subespacio propio para T sobre el correspondiente subespacio para T'. En efecto, por (9-12), se tiene que

$$T' = \sum_j c_j E'_j, \qquad E'_j = UE_jU^{-1}$$

es la descomposición espectral de T'. Luego

$$\begin{aligned} f(T') &= \sum_j f(c_j)E'_j \\ &= \sum_j f(c_j)UE_jU^{-1} \\ &= U\left(\sum_j f(c_j)E_j\right)U^{-1} \\ &= Uf(T)U^{-1}. \quad \blacksquare \end{aligned}$$

Pensando en el análisis anterior, es importante tener presente que el espectro del operador normal T es el conjunto

$$S = \{c_1, \ldots, c_k\}$$

de valores propios distintos. Cuando T está representado por una matriz diagonal en una base de vectores propios es necesario repetir cada valor c_j tantas veces como la dimensión del correspondiente espacio de vectores propios. Esta es la razón del cambio de notación en el siguiente resultado.

Corolario. *Con las hipótesis del Teorema* 10, *supóngase que T está representado en la base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ por la matriz diagonal D de elementos $d_1, \ldots, d_n$. Entonces, en la base $\mathcal{B}$, $f(T)$ está representada por la matriz diagonal $f(D)$ de elementos $f(d_1), \ldots, f(d_n)$. Si $\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ es otra base ordenada cualquiera y P es la matriz tal que*

$$\alpha'_j = \sum_i P_{ij}\alpha_i$$

entonces $P^{-1}f(D)P$ es la matriz de $f(T)$ en la base $\mathcal{B}'$.

Demostración. Para cada índice i existe un único j tal que $1 \leq j \leq k$, α_i pertenece a $E_j(V)$ y $d_i = c_j$. Luego $f(T)\alpha_i = f(d_i)\alpha_i$, para todo i, y

$$\begin{aligned} f(T)\alpha'_j &= \sum_i P_{ij}f(T)\alpha_i \\ &= \sum_i d_iP_{ij}\alpha_i \\ &= \sum_i (DP)_{ij}\alpha_i \\ &= \sum_i (DP)_{ij} \sum_k P^{-1}_{ki}\alpha'_k \\ &= \sum_k (P^{-1}DP)_{kj}\alpha'_k. \quad \blacksquare \end{aligned}$$

Se sigue de este resultado que pueden formarse ciertas funciones de una matriz normal. Pues supóngase que A es una matriz normal. Entonces existe una matriz inversible P, en realidad una matriz unitaria P, tal que PAP^{-1} es diagonal, sea D con elementos $d_1, \ldots, d_n$. Sea f una función compleja que se pueda aplicar a $d_1, \ldots, d_n$ y sea $f(D)$ la matriz diagonal con elementos $f(d_1), \ldots, f(d_n)$. Entonces $P^{-1}f(D)P$ es independiente de D y es una función de A en el siguiente sentido: Si Q es otra matriz inversible tal que QAQ^{-1} sea una matriz diagonal D', entonces f puede aplicarse a los elementos de la diagonal de D', y

$$P^{-1}f(D)P = Q^{-1}f(D')Q.$$

Definición. *En las condiciones anteriores, $f(A)$ se define como $P^{-1}f(D)P$.*

La matriz $f(A)$ puede ser también caracterizada de un modo distinto. Al hacerlo se enuncian sin demostración algunos de los resultados sobre matrices normales que se obtienen al formular los teoremas con matrices análogos de los teoremas anteriores.

Teorema 11. *Sean A una matriz normal y $c_1, \ldots, c_k$ las distintas raíces complejas de $\det(xI - A)$. Sea*

$$e_i = \prod_{j \neq i} \left(\frac{x - c_j}{c_i - c_j} \right)$$

y $E_i = e_i(A)$ $(1 \leq i \leq k)$. Entonces $E_iE_j = 0$ cuando $i \neq j$, $E_i^2 = E_i$, $E_i^* = E_i$, y

$$I = E_1 + \cdots + E_k.$$

Si f es una función compleja cuyo dominio incluye a $c_1, \ldots, c_k$, entonces

$$f(A) = f(c_1)E_1 + \cdots + f(c_k)E_k;$$

en particular, $A = c_1E_1 + \cdots + c_kE_k$.

Recordemos que un operador en un espacio producto interior V es no negativo si T es autoadjunto y $(T\alpha|\alpha) \geq 0$ para todo α de V.

Teorema 12. *Sea T un operador diagonalizable normal sobre un espacio producto interno de dimensión finita V. Entonces T es autoadjunto, no negativo, o unitario, según que cada valor característico de T sea real, no negativo, o de valor absoluto 1.*

Demostración. Supóngase que T tenga descomposición espectral $T = c_1E_1 + \cdots + c_kE_k$, entonces $T^* = \bar{c}_1E_1 + \cdots + \bar{c}_kE_k$. Decir que T es autoadjunto es decir que $T = T^*$, o

$$(c_1 - \bar{c}_1)E_1 + \cdots + (c_k - \bar{c}_k)E_k = 0.$$

Valiéndose de que $E_iE_j = 0$ para $i \neq j$ y de que ningún E_j es el operador cero,

se ve que T es autoadjunto si, y solo si, $c_j = \bar{c}_j$, $j = 1, \ldots, k$. Para distinguir los operadores normales que son no negativos, obsérvese que

$$\begin{aligned}(T\alpha|\alpha) &= \left(\sum_{j=1}^{k} c_jE_j\alpha \middle| \sum_{i=1}^{k} E_i\alpha\right)\\ &= \sum_i \sum_j c_j(E_j\alpha|E_i\alpha)\\ &= \sum c_j||E_j\alpha||^2.\end{aligned}$$

Hemos usado el hecho de que $(E_j\alpha|E_i\alpha) = 0$ para $i \neq j$. De esto resulta que la condición $(T\alpha|\alpha) \geq 0$ se satisface si, y solo si, $c_j \geq 0$ para todo j. Para distinguir los operadores unitarios obsérvese que

$$\begin{aligned}TT^* &= c_1c_1E_1 + \cdots + c_kc_kE_k\\ &= |c_1|^2E_1 + \cdots + |c_k|^2E_k.\end{aligned}$$

Si $TT^* = I$, entonces $I = |c_1|^2E_1 + \cdots + |c_k|^2E_k$, y operando con E_j,

$$E_j = |c_j|^2E_j.$$

Como $E_j \neq 0$, se tiene que $|c_j|^2 = 1$, o $|c_j| = 1$. Recíprocamente, si $|c_j|^2 = 1$ para cada j, es evidente que $TT^* = I$. ∎

Es importante hacer notar que éste es un teorema sobre operadores normales. Si T es un operador lineal general sobre V, que tiene valores propios reales, no se sigue que T sea autoadjunto. El teorema dice que si T tiene valores propios reales y si T es diagonalizable y normal, entonces T es autoadjunto. Un teorema de este tipo sirve para reforzar la analogía entre la operación de adjunción y el proceso de formar el conjugado de un número complejo. Un número complejo z es real o de valor absoluto 1 según que $z = \bar{z}$, o $z\bar{z} = 1$. Un operador T es autoadjunto o unitario según que $T = T^*$ o $TT^* = I$.

Se demostrarán ahora dos teoremas análogos de las dos afirmaciones siguientes:

(1) Todo número no negativo tiene una única raíz cuadrada no negativa.

(2) Todo número complejo se puede expresar en la forma ru, con r no negativo y $|u| = 1$. Esta es la descomposición polar $z = re^{i\theta}$ de los números complejos.

Teorema 13. *Sean V un espacio producto interno de dimensión finita y T un operador no negativo sobre V. Entonces T tiene una única raíz cuadrada no negativa, esto es, existe un operador no negativo N, y solo uno, de V tal que $N^2 = T$.*

Demostración. Sea $T = c_1E_1 + \cdots + c_kE_k$ la descomposición espectral de T. Por el Teorema 12, todo $c_j \geq 0$. Si c es cualquier número real no negativo, sea $\sqrt{c}$ la raíz cuadrada no negativa de c. Entonces, conforme al Teorema 11 y a (9-12), $N = \sqrt{T}$ es un operador diagonalizable normal bien definido sobre V. Es no negativo por el Teorema 12 y, por un cálculo obvio, $N^2 = T$.

Ahora sea P un operador no negativo sobre V tal que $P^2 = T$. Sea demostrar que $N = P$. Sea

$$P = d_1F_1 + \cdots + d_rF_r$$

la descomposición espectral de P. Entonces $d_j \geq 0$ para todo j, pues P es no negativo. De $P^2 = T$ se tiene

$$T = d_1^2F_1 + \cdots + d_r^2F_r.$$

Ahora $F_1, \ldots, F_r$ cumple las condiciones $I = F_1 + \cdots + F_r$, $F_iF_j = 0$ para $i \neq j$, y ningún F_j es 0. Los números $d_1^2, \ldots, d_r^2$ son distintos, ya que números no negativos distintos tienen raíces cuadradas distintas. Por la unicidad de la descomposición espectral de T, se tiene $r = k$ y (si acaso por reordenación) $F_j = E_j$, $d_j^2 = c_j$. Así que $P = N$. ∎

Teorema 14. *Sea V un espacio producto interno de dimensión finita y sea T un operador lineal cualquiera sobre V. Entonces existen un operador unitario U sobre V y un operador no negativo N sobre V tales que $T = UN$. El operador N no negativo es único. Si T es inversible, el operador U es también único.*

Demostración. Supóngase que $T = UN$, donde U es unitario y N es no negativo. Entonces $T^* = (UN)^* = N^*U^* = NU^*$. Así $T^*T = NU^*UN = N^2$. Esto muestra que N está unívocamente determinado como la raíz cuadrada no negativa del operador no negativo T^*T.

De modo que para comenzar con la demostración de la existencia de U y de N, se usa el Teorema 13 para definir N como la única raíz cuadrada no negativa de T^*T. Si T es inversible, también lo es N, pues

$$(N\alpha|N\alpha) = (N^2\alpha|\alpha) = (T^*T\alpha|\alpha) = (T\alpha|T\alpha).$$

En este caso se define $U = TN^{-1}$ y se demuestra que U es unitario. Ahora bien, $U^* = (TN^{-1})^* = (N^{-1})^*T^* = (N^*)^{-1}T^* = N^{-1}T^*$. Así que

$$\begin{aligned} UU^* &= TN^{-1}N^{-1}T^* \\ &= T(N^{-1})^2T^* \\ &= T(N^2)^{-1}T^* \\ &= T(T^*T)^{-1}T^* \\ &= TT^{-1}(T^*)^{-1}T^* \\ &= I \end{aligned}$$

y U es unitario.

Si T no es inversible hay que proceder en forma más laboriosa para definir U. Primeramente se define U sobre la imagen de N. Si α es un vector de esta imagen, por ejemplo, $\alpha = N\beta$. Se define $U\alpha = T\beta$, motivado por el hecho de que se quiere que $UN\beta = T\beta$. Debe verificarse que U está bien definido sobre la imagen de N; es decir, si $N\beta' = N\beta$, entonces $T\beta' = T\beta$. Se verificó anteriormente que $\|N\gamma\|^2 = \|T\gamma\|^2$ para todo γ en V. Así, con $\gamma = \beta - \beta'$ se ve que $N(\beta - \beta') = 0$ si, y solo si, $T(\beta = \beta') = 0$. Con lo que U está bien definido sobre la imagen de N y es evidentemente lineal donde está definido. Ahora bien, si W es la imagen de N, se definirá U sobre $W^\perp$. Para ello se necesita la

siguiente observación. Como T y N tienen el mismo espacio nulo, sus imágenes tienen la misma dimensión. Así $W^{\perp}$ tiene la misma dimensión que el complemento ortogonal de la imagen de T. Por tanto, existe un isomorfismo (de espacio producto interno) U_0 de $W^{\perp}$ sobre $T(V)^{\perp}$. Ahora se ha definido U sobre W, y se define U sobre $W^{\perp}$ como U_0.

Se repite la definición de U. Como $V = W \oplus W^{\perp}$, todo α de V está expresado unívocamente en la forma $\alpha = N\beta + \gamma$, donde $N\beta$ está en la imagen W de N y γ está en $W^{\perp}$. Se define

$$U\alpha = T\beta + U_0\gamma.$$

Este U es evidentemente lineal, y se verificó anteriormente que está bien definido. Así, pues,

$$\begin{aligned}(U\alpha|U\alpha) &= (T\beta + U_0\gamma|T\beta + U_0\gamma)\\ &= (T\beta|T\beta) + (U_0\gamma|U_0\gamma)\\ &= (N\beta|N\beta) + (\gamma|\gamma)\\ &= (\alpha|\alpha)\end{aligned}$$

y entonces U es unitario. También tenemos que $UN\beta = T\beta$ para todo β. ∎

$T = UN$ se le llama una **descomposición polar** de T. Por cierto que no se le puede llamar *la* descomposición polar, ya que U no es único. Aun cuando T sea inversible, con lo que U es único, tenemos la dificultad de que U y N pueden no conmutar. Ciertamente, conmutan si, y solo si, T es normal. Por ejemplo, si $T = UN = NU$, con N no negativo y U unitario, entonces

$$TT^* = (NU)(NU)^* = NUU^*N = N^2 = T^*T.$$

El operador general T tendrá también una descomposición $T = N_1U_1$ con N_1 no negativo y U_1 unitario. Aquí N_1 será la raíz cuadrada no negativa de TT^*. Se puede obtener este resultado aplicando el teorema que se acaba de demostrar al operador T^*, y tomando luego adjuntos.

Pasamos ahora al problema de qué se puede decir respecto a la diagonalización simultánea de familias conmutativas de operadores normales. Para este propósito es adecuada la siguiente terminología.

Definiciones. *Sea $\mathfrak{F}$ una familia de operadores sobre un espacio producto interno V. Una función r sobre $\mathfrak{F}$, con valores en el cuerpo F de los escalares, se llamará una* **raíz** *de $\mathfrak{F}$ si existe un α no nulo de V tal que*

$$T\alpha = r(T)\alpha$$

para todo T de $\mathfrak{F}$. Para cualquier función r de $\mathfrak{F}$ en F, sea $V(r)$ el conjunto de todos los α de V tales que $T\alpha = r(T)\alpha$ para todo T en $\mathfrak{F}$.

Entonces $V(r)$ es un subespacio de V y r es una raíz de $\mathfrak{F}$ si, y solo si, $V(r) \neq 0$. Todo α no nulo de $V(r)$ es simultáneamente un vector propio para todo T de $\mathfrak{F}$.

Teorema 15. *Sea $\mathfrak{F}$ una familia conmutativa de operadores normales diagonalizables en un espacio producto interno de dimensión finita V. Entonces $\mathfrak{F}$ tiene*

solo un número finito de raíces. Si $r_1, \ldots, r_k$ son las raíces distintas de $\mathfrak{F}$, entonces

(i) *$V(r_i)$ es ortogonal a $V(r_j)$ cuando $i \neq j$, y*
(ii) *$V = V(r_1) \oplus \cdots \oplus V(r_k)$.*

Demostración. Supóngase que r y s son raíces distintas de $\mathfrak{F}$. Entonces existe un operador T en $\mathfrak{F}$ tal que $r(T) \neq s(T)$. Como los vectores propios que corresponden a valores propios distintos de T son necesariamente ortogonales, se sigue que $V(r)$ es ortogonal a $V(s)$. Como V es de dimensión finita, esto implica que $\mathfrak{F}$ tiene a lo más un número finito de raíces. Sean $r_1, \ldots, r_k$ las raíces de $\mathfrak{F}$. Supóngase que $\{T_1, \ldots, T_m\}$ es un subconjunto maximal de $\mathfrak{F}$ linealmente independiente y sea

$$\{E_{i1}, E_{i2}, \ldots\}$$

la descomposición de la identidad definida por T_i $(1 \leq i \leq m)$. Entonces las proyecciones E_{ij} forman una familia conmutativa, pues toda E_{ij} es un polinomio en T_i y $T_1, \ldots, T_m$ conmutan entre sí. Como

$$I = (\sum_{j_1} E_{1j_1}) (\sum_{j_2} E_{2j_2}) \cdots (\sum_{j_m} E_{mj_m})$$

todo vector α de V puede escribirse en la forma

$$\alpha = \sum_{j_1, \ldots, j_m} E_{1j_1}E_{2j_2} \cdots E_{mj_m}\alpha. \tag{9-13}$$

Supóngase que $j_1, \ldots, j_m$ son los índices para los que $\beta = E_{1j_1}E_{2j_2} \cdots E_{mj_m} \neq 0$. Sea

$$\beta_i = (\prod_{n \neq i} E_{nj_n})\,\alpha.$$

Entonces $\beta = E_{ij}\beta_i$; luego existe un escalar c_i tal que

$$T_i\beta = c_i\beta, \qquad 1 \leq i \leq m.$$

Para cada T en $\mathfrak{F}$ existen escalares únicos b_i tales que

$$T = \sum_{i=1}^{m} b_iT_i.$$

Así

$$T\beta = \sum_i b_iT_i\beta$$

$$= (\sum_i b_ic_i)\,\beta.$$

La función $T \rightarrow \sum_i b_ic_i$ es evidentemente una de las raíces, por ejemplo, r_t de $\mathfrak{F}$, y β está en $V(r_t)$. Por tanto, cada término no nulo en (9-13) pertenece a uno de los espacios $V(r_1), \ldots, V(r_k)$. Se sigue que V es la suma directa ortogonal de $V(r_1), \ldots, V(r_k)$. ∎

Corolario. *En la hipótesis del teorema, sea P_j la proyección ortogonal de V sobre $V(r_j)$, $(1 \leq j \leq k)$. Entonces $P_iP_j = 0$ si $i \neq j$.*

$$I = P_1 + \cdots + P_k,$$

y todo T de $\mathcal{F}$ *puede ser escrito en la forma*

$$T = \sum_j r_j(T)P_j. \tag{9-14}$$

Definiciones. *La familia de las proyecciones ortogonales* $\{P_1, \ldots, P_k\}$ *se llama* **descomposición de la identidad determinada por** $\mathcal{F}$ *y* (9-14) *es la* **descomposición espectral de** T **según esta familia.**

Aunque las proyecciones $P_1, \ldots, P_k$ del corolario anterior están asociadas canónicamente con la familia $\mathcal{F}$, no están generalmente en $\mathcal{F}$, ni son siquiera combinaciones lineales de operadores de $\mathcal{F}$; sin embargo, veremos que se pueden obtener por la formación de ciertos productos de polinomios en elementos de $\mathcal{F}$.

Para el estudio de una familia de operadores en un espacio producto interno suele ser provechoso considerar el álgebra autoadjunta generada por la familia.

Definición. *Un* **álgebra autoadjunta de operadores** *sobre un espacio producto interno V es un subálgebra lineal de L(V, V) que contiene el adjunto de cada uno de sus elementos.*

Ejemplo de álgebra autoadjunta es el mismo $L(V, V)$. Como la intersección de cualquier colección de álgebras autoadjuntas es a su vez un álgebra autoadjunta, la siguiente terminología tiene sentido.

Definición. *Si* $\mathcal{F}$ *es una familia de operadores lineales sobre un espacio producto interno de dimensión finita, el* **álgebra autoadjunta generada por** $\mathcal{F}$ *es la menor álgebra autoadjunta que contiene a* $\mathcal{F}$.

Teorema 16. *Sea* $\mathcal{F}$ *una familia conmutativa de operadores diagonalizables normales sobre un espacio producto interno de dimensión finita V y sea* $\mathcal{A}$ *el álgebra autoadjunta generada por* $\mathcal{F}$ *y el operador identidad. Sea* $\{P_1, \ldots, P_k\}$ *la descomposición de la identidad definida por* $\mathcal{F}$. *Entonces* $\mathcal{A}$ *es el conjunto de todos los operadores sobre V de la forma*

$$T = \sum_{j=1}^{k} c_jP_j \tag{9-15}$$

donde $c_1, \ldots, c_k$ *son escalares arbitrarios.*

Demostración. Sea $\mathcal{C}$ el conjunto de todos los operadores sobre V de la forma (9-15). Entonces $\mathcal{C}$ contiene al operador identidad y el adjunto

$$T^* = \sum_j \bar{c}_jP_j$$

para cada uno de sus elementos. Si $T = \sum_j c_jP_j$ y $U = \sum_j d_jP_j$, entonces para todo escalar a

$$aT + U = \sum_j (ac + d_j)P_j$$

y

$$TU = \sum_{i,j} c_i d_j P_i P_j$$
$$= \sum_j c_j d_j P_j$$
$$= UT.$$

Así que $\mathcal{C}$ es un álgebra autoadjunta conmutativa que contiene a $\mathcal{F}$ y al operador identidad. Por tanto, $\mathcal{C}$ contiene a $\mathcal{A}$.

Ahora sean $r_1, \ldots, r_k$ todas las raíces de $\mathcal{F}$. Entonces, para todo par de índices (i, n) con $i \neq n$, existe un operador T_{in} en $\mathcal{F}$ tal que $r_i(T_{in}) \neq r_n(T_{in})$. Sea $a_{in} = r_i(T_{in}) - r_n(T_{in})$ y $b_{in} = r_n(T_{in})$. Entonces el operador lineal

$$Q_i = \prod_{n \neq i} a_{in}^{-1}(T_{in} - b_{in}I)$$

es un elemento del álgebra $\mathcal{A}$. Se demuestra que $Q_i = P_i$ $(1 \leq i \leq k)$. Para ello, supóngase que $j \neq i$ y que α es un vector arbitrario en $V(r_j)$. Entonces

$$T_{ij}\alpha = r_j(T_{ij})\alpha$$
$$= b_{ij}\alpha$$

de modo que $(T_{ij} = b_{ij}I)\alpha = 0$. Como todos los factores en Q_i conmutan, se sigue que $Q_i\alpha = 0$. Luego Q_i coincide con P_i en $V(r_j)$ si $j \neq i$. Supóngase ahora que α es un vector de $V(r_i)$. Entonces $T_{in}\alpha = r_i(T_{in})\alpha$, y

$$a_{in}^{-1}(T_{in} - b_{in}I)\alpha = a_{in}^{-1}[r_i(T_{in}) - r_n(T_{in})]\alpha = \alpha.$$

Así, pues, $Q_i\alpha = \alpha$ y Q_i coincide con P_i en $V(r_i)$; por tanto, $Q_i = P_i$ para i $1, \ldots, k$. De lo cual se sigue que $\mathcal{A} = \mathcal{C}$. ▌

El teorema muestra que el álgebra $\mathcal{A}$ es conmutativa y que cada elemento de $\mathcal{A}$ es un operador normal diagonalizable. Se demuestra a continuación que $\mathcal{A}$ tiene un solo generador.

Corolario. *Con las hipótesis del teorema existe un operador T en $\mathcal{A}$ tal que todo elemento de $\mathcal{A}$ es un polinomio en T.*

Demostración. Sea $= \sum_{j=1}^{k} t_j P_j$ con $t_1, \ldots, t_k$ son escalares distintos. Entonces

$$T^n = \sum_{j=1}^{k} t_j^n P_j$$

para $n = 1, 2, \ldots$ Si

$$f = \sum_{n=1}^{s} a_n x^n$$

se sigue que

$$f(T) = \sum_{n=1}^{s} a_n T^n = \sum_{n=1}^{s} \sum_{j=1}^{k} a_n t_j^n P_j$$
$$= \sum_{j=1}^{k} \left(\sum_{n=1}^{s} a_n t_j^n \right) P_j$$
$$= \sum_{j=1}^{k} f(t_j) P_j.$$

Dado un

$$U = \sum_{j=1}^{k} c_j P_j$$

arbitrario en $\mathcal{Q}$, existe un polinomio f tal que $f(t_j) = c_j$ $(1 \leq j \leq k)$ y, para cualquier tal f, $U = f(T)$. ▌

Ejercicios

1. Dar una definición razonable de una matriz $n \times n$ no negativa y demostrar luego que tal matriz tiene una única raíz cuadrada no negativa.

2. Sea A una matriz $n \times n$ con elementos complejos tal que $A^* = -A$ y sea $B = e^A$. Demostrar que

(a) $\det B = e^{\operatorname{tr} A}$;
(b) $B^* = e^{-A}$;
(c) B es unitaria.

3. Si U y T son operadores normales que conmutan, demostrar que $U + T$ y UT son normales.

4. Sea T un operador lineal sobre un espacio producto interno de dimensión finita V. Demostrar que las diez siguientes afirmaciones respecto a T son equivalentes:
(a) T es normal.
(b) $\|T\alpha\| = \|T^*\alpha\|$ para todo α en V.
(c) $T = T_1 + iT_2$, donde T_1 y T_2 son autoadjuntos y $T_1 T_2 = T_2 T_1$.
(d) Si α es un vector y c un escalar de modo que $T\alpha = c\alpha$, entonces $T^*\alpha = \bar{c}\alpha$.
(e) Existe una base ortonormal de V formada por vectores propios para T.
(f) Existe una base ortonormal $\mathcal{B}$ tal que $[T]_{\mathcal{B}}$ es diagonal.
(g) Existe un polinomio g con coeficientes complejos tal que $T^* = g(T)$.
(h) Todo subespacio invariante por T es también invariante por T^*.
(i) $T = NU$, donde N es no negativo, U es unitario, y N conmuta con U.
(j) $T = c_1 T_1 + \cdots + c_k E_k$ donde $I = E_1 + \cdots + E_k$, $E_i E_j = 0$ para $i \neq j$, y $E_j^2 = E_j = E_j^*$.

5. Usar el Ejercicio 3 para mostrar que cualquier familia conmutativa de operadores normales (no necesariamente diagonalizables) en un espacio producto interno de dimensión finita genera un álgebra autoadjunta conmutativa de operadores normales.

6. Sean V un espacio producto interno complejo de dimensión finita y U un operador unitario sobre V tal que $U\alpha = \alpha$ implique $\alpha = 0$. Sea

$$f(z) = i\frac{(1+z)}{(1-z)}, \qquad z \neq 1$$

y demostrar que
(a) $f(U) = i(I + U)(I - U)^{-1}$.
(b) $f(U)$ es autoadjunto.
(c) para todo operador autoadjunto T sobre V, el operador

$$U = (T - iI)(T + iI)^{-1}$$

es unitario y tal que $T = f(U)$.

7. Sea V el espacio de las matrices complejas $n \times n$ con el producto interno

$$(A|B) = \operatorname{tr}(AB^*).$$

Si B es un elemento de V, sean L_B, R_B y T_B los operadores lineales sobre V definidos por

(a) $L_B(A) = BA$.
(b) $R_B(A) = AB$.
(c) $T_B(A) = BA - AB$.

Considerar las tres familias de operadores que se obtienen al hacer variar B sobre todas las matrices diagonales. Demostrar que cada una de estas familias es un álgebra autoadjunta conmutativa y hallar sus descomposiciones espectrales.

8. Si B es un elemento del espacio producto interno del Ejercicio 7, hacer ver que L_B es equivalente unitariamente a R_B.

9. Sean V el espacio producto interno del Ejercicio 7 y G el grupo de matrices unitarias de V. Si B está en G, sea C_B el operador lineal sobre V definido por

$$C_B(A) = BAB^{-1}.$$

Demostrar que

(a) C_B es un operador unitario en V;
(b) $C_{B_1B_2} = C_{B_1}C_{B_2}$;
(c) no existe una transformación unitaria U en V tal que

$$UL_BU^{-1} = C_B$$

para todo B de G.

10. Sean $\mathfrak{F}$ una familia cualquiera de operadores lineales sobre un espacio producto interno de dimensión finita V y $\mathcal{A}$ el álgebra autoadjunta generada por $\mathfrak{F}$. Demostrar que

(a) toda raíz de $\mathcal{A}$ define una raíz de $\mathfrak{F}$;
(b) toda raíz r de $\mathcal{A}$ es una función lineal multiplicativa sobre A, es decir,

$$r(TU) = r(T)r(U)$$
$$r(cT + U) = cr(T) + r(U)$$

para todo T y U de $\mathcal{A}$ y todo escalar c.

11. Sea $\mathfrak{F}$ una familia conmutativa de operadores diagonalizables normales sobre un espacio producto interno de dimensión finita V y sea $\mathcal{A}$ el álgebra autoadjunta generada por $\mathfrak{F}$ y el operador identidad I. Demostrar que toda raíz de $\mathcal{A}$ es distinta de 0 y que para toda raíz r de $\mathfrak{F}$ existe una raíz única s de $\mathcal{A}$ tal que $s(T) = r(T)$ para todo T de $\mathfrak{F}$.

12. Sean $\mathfrak{F}$ una familia conmutativa de operadores diagonalizables normales en un espacio producto interno de dimensión finita V y $\mathcal{A}_0$ el álgebra autoadjunta generada por $\mathfrak{F}$. Sea $\mathcal{A}$ el álgebra autoadjunta generada por $\mathfrak{F}$ y el operador identidad I. Demostrar que

(a) $\mathcal{A}$ es el conjunto de todos los operadores sobre V de la forma $cI + T$, donde c es un escalar y T un operador en $\mathcal{A}_0$.

(b) Existe a lo más una raíz r de $\mathcal{A}$ tal que $r(T) = 0$ para todo T de $\mathcal{A}_0$.

(c) Si una de las raíces de $\mathcal{A}$ es 0 sobre $\mathcal{A}_0$, las proyecciones $P_1, \ldots, P_k$ en la descomposición de la identidad definida por $\mathfrak{F}$ pueden ser rotuladas mediante índices de tal modo que $\mathcal{A}_0$ conste de todos los operadores sobre V de la forma

$$T = \sum_{j=2}^{k} c_jP_j$$

donde $c_2, \ldots, c_k$ son escalares arbitrarios.

(d) $\mathcal{A} = \mathcal{A}_0$ si, y solo si, para cada raíz r de $\mathcal{A}$ existe un operador T en $\mathcal{A}_0$ tal que $r(T) \neq 0$.

9.6. Otras propiedades de los operadores normales

En la Sección 8.5 se han expuesto las propiedades básicas de los operadores autoadjuntos y normales usando el método más simple y directo posible. En la Sección 9.5 se consideraron varios aspectos de la teoría espectral. Aquí demostraremos algunos resultados de naturaleza más técnica y que se refieren principalmente a operadores normales sobre espacios reales.

Se comenzará demostrando una versión más precisa del teorema de descomposición prima del Capítulo 6 para operadores normales. Será válido para ambos casos el real y el complejo.

Teorema 17. *Sea T un operador normal sobre un espacio producto interno de dimensión finita V. Sea p el polinomio minimal para T y $p_1, \ldots, p_k$ sus distintos factores primos mónicos. Entonces cada p_j tiene multiplicidad 1 en la factorización de p y tiene grado 1 o 2. Supóngase que W_j es el espacio nulo de $p_j(T)$. Entonces*

(i) *W_j es ortogonal a W_i, si $i \neq j$;*
(ii) *$V = W_1 \oplus \cdots \oplus W_k$;*
(iii) *W_j es invariante por T y p_j es el polinomio minimal para la restricción de T a W_j;*
(iv) *para todo j existe un polinomio e_j con coeficientes en el cuerpo de los escalares tal que $e_j(T)$ es la proyección ortogonal de V sobre W_j.*

En la demostración usaremos ciertos hechos básicos que sentaremos como lemas.

Lema 1. *Sea N un operador normal sobre un espacio producto interno W. Entonces el espacio nulo de N es el complemento ortogonal de su imagen.*

Demostración. Supóngase que $(\alpha|N\beta) = 0$ para todo β de W. Entonces $(N^*\alpha|\beta) = 0$ para todo β; luego $N^*\alpha = 0$. Por el Teorema 19 del Capítulo 8 esto implica que $N\alpha = 0$. Recíprocamente, si $N\alpha = 0$ entonces $N^*\alpha = 0$ y

$$(N^*\alpha|\beta) = (\alpha|N\beta) = 0$$

para todo β de W. ▌

Lema 2. *Si N es un operador normal y α es un vector tal que $N^2\alpha = 0$, entonces $N\alpha = 0$.*

Demostración. Supóngase que N sea normal y que $N^2\alpha = 0$. Entonces $N\alpha$ está en la imagen de N y también en el espacio nulo de N. Por el Lema 1, esto implica que $N\alpha = 0$. ▌

Lema 3. *Sean T un operador normal y f cualquier polinomio con coeficientes en el cuerpo de los escalares. Entonces $f(T)$ es también normal.*

Demostración. Supóngase que $f = a_0 + a_1x + \cdots + a_nx^n$. Entonces

$$f(T) = a_0I + a_1T + \cdots + a_nT^n$$

$$f(T)^* = \bar{a}_0I + \bar{a}_1T^* + \cdots + \bar{a}_n(T^*)^n.$$

Como $T^*T = TT^*$, se sigue que $f(T)$ conmuta con $f(T)^*$. ▌

Lema 4. *Sean T un operador normal y f, g polinomios primos relativos con coeficientes en el cuerpo de los escalares. Supóngase que α y β son vectores tales que $f(T)\alpha = 0$ y $g(T)\beta = 0$. Entonces $(\alpha|\beta) = 0$.*

Demostración. Existen polinomios a y b con coeficientes en el cuerpo de los escalares tales que $af + bg = 1$. Así

$$a(T)f(T) + b(T)g(T) = I$$

y $\alpha = g(T)b(T)\alpha$. Se sigue que

$$(\alpha|\beta) = (g(T)b(T)\alpha|\beta) = (b(T)\alpha|g(T)^*\beta).$$

Por suposición, $g(T)\beta = 0$. Por el Lema 3, $g(T)$ es normal. Por tanto, por el Teorema 19 del Capítulo 8, $g(T)^*\beta = 0$; luego $(\alpha|\beta) = 0$. ▌

Demostración del Teorema 17. Se recuerda que el polinomio minimal para T es el polinomio mónico de grado menor entre todos los polinomios f tales que $f(T) = 0$. La existencia de tales polinomios se desprende de la suposición de que V es de dimensión finita. Supóngase que un factor primo p_j de p está repetido. Entonces $p = p_j^2g$ para algún polinomio g. Como $p(T) = 0$, se sigue que

$$(p_j(T))^2g(T)\alpha = 0$$

para todo α de V. Por el Lema 3, $p_j(T)$ es normal. Así, el Lema 2 implica que

$$p_j(T)g(T)\alpha = 0$$

para todo α de V. Pero esto contradice la suposición de que p tiene el menor grado entre todos los f tales que $f(T) = 0$. Por tanto, $p = p_1 \cdots p_k$. Si V es un espacio producto interno complejo, todo p_j es necesariamente de la forma

$$p_j = x - c_j$$

con c_j real o complejo. Por otro lado, si V es un espacio producto interno real, entonces $p_j = x_j - c_j$ con c_j en R o

$$p_j = (x - c)(x - \bar{c})$$

donde c es un número complejo no real.

Sea ahora $f_j = p/p_j$. Entonces, como $f_1, \ldots, f_k$ son primos relativos, existen polinomios g_j con coeficientes en el cuerpo de los escalares tales que

(9-16) $$1 = \sum_j f_j g_j.$$

Indicamos brevemente cómo se pueden construir estos g_j. Si $p_j = x - c_j$, entonces $f_j(c_j) \neq 0$, y para g_j se toma el polinomio escalar $1/f_j(c_j)$. Si todo polinomio es de esta forma, los $f_j g_j$ son la familia de los polinomios de Lagrange asociados con los $c_1, \ldots, c_k$, y (9-16) es claramente válida. Supóngase que algún $p_j = (x - c)(x - \bar{c})$ con c un número complejo no real. Entonces V es un espacio producto interno real, y se tiene

$$g_j = \frac{x - \bar{c}}{s} + \frac{x - c}{\bar{s}}$$

donde $s = (c - \bar{c})f_j(c)$. Entonces

$$g_j = \frac{(s + \bar{s})x - (cs + \bar{c}s)}{s\bar{s}}$$

con lo que g_j es un polinomio con coeficientes reales. Si p es de grado n, entonces

$$1 - \sum_j f_j g_j$$

es un polinomio con coeficientes reales de grado $n - 1$ a lo más; además, se anula en cada una de las n raíces (complejas) de p' y, por tanto, es idénticamente 0.

Sea ahora α un vector arbitrario de V. Entonces, por (9-16)

$$\alpha = \sum_j f_j(T)g_j(T)\alpha$$

y como $p_j(T)f_j(T) = 0$, se sigue que $f_j(T)g_j(T)\alpha$ está en W_j para cada j. Por el Lema 4, W_j es ortogonal a W_i siempre que $i \neq j$. Por tanto, V es la suma directa ortogonal de $W_1, \ldots, W_k$. Si β es cualquier vector en W_j, entonces

$$p_j(T)T\beta = Tp_j(T)\beta = 0;$$

así que W_j es invariante por T. Sea T_j la restricción de T a W_j. Entonces $p_j(T_j) = 0$, con lo que p_j es divisible por el polinomio minimal para T_j. Como p_j es irreducible sobre el cuerpo de los escalares, se sigue que p_j es el polinomio minimal para T_j.

Sea ahora $e_j = f_j g_j$ y $E_j = e_j(T)$. Entonces para todo vector α de V $E_j\alpha$ está en W_j, y

$$\alpha = \sum_j E_j\alpha.$$

Así, $\alpha - E_i\alpha = \sum_{j \neq i} E_j\alpha$; como W_j es ortogonal a W_i si $j \neq i$, esto implica que $\alpha - E_i\alpha$ está en $W_i^{\perp}$. Se deduce ahora, del Teorema 4 del Capítulo 8, que E_i es la proyección ortogonal de V sobre W_i. ▌

Definición. *Los subespacios* W_j $(1 \leq j \leq k)$ *se llaman* **componentes primos de** V **según** T.

Corolario. *Sean T un operador normal sobre un espacio producto interno de dimensión finita V y $W_1, \ldots, W_k$ los componentes primos de V según T. Supóngase que W es un subespacio de V invariante por T. Entonces*

$$W = \sum_j W \cap W_j.$$

Demostración. Evidentemente W contiene a $\sum_j W \cap W_j$. Por otro lado, W, que es invariante por T, es invariante por todo polinomio de T. En particular, W es invariante por la proyección ortogonal E_j de V sobre W_j. Si α está en W, se sigue que $E_j\alpha$ está en $W \cap W_j$ y, al mismo tiempo, que $\alpha = \sum_j E_j\alpha$. Por tanto, W está contenido en $\sum_j W \cap W_j$. ∎

El Teorema 17 muestra que todo operador normal T sobre un espacio producto interno de dimensión finita está dado canónicamente por un número finito de operadores normales T_j definidos sobre los componentes primos W_j de V según T, cada uno de cuyos polinomios minimales es irreducible sobre el cuerpo de los escalares. Para completar el conocimiento de los operadores normales es necesario estudiar operadores normales de este tipo especial.

Un operador normal cuyo polinomio minimal es de grado 1 es obviamente apenas un múltiplo escalar de la identidad. Por otro lado, cuando el polinomio minimal es irreducible y de grado 2, la situación es más complicada.

Ejemplo 1. Supóngase que $r > 0$ y que θ sea un número real no múltiplo entero de π. Sea T el operador lineal sobre R^2 cuya matriz en la base ortonormal canónica es

$$A = r\begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

Entonces T es un múltiplo escalar de una transformación ortogonal y, por tanto, normal. Sea p el polinomio propio de T, entonces

$$\begin{aligned} p &= \det(xI - A) \\ &= (x - r\cos\theta)^2 + r^2\operatorname{sen}^2\theta \\ &= x - 2r\cos\theta x + r^2. \end{aligned}$$

Sea $a = r\cos\theta$, $b = r\operatorname{sen}\theta$ y $c = a + ib$. Entonces $b \neq 0$, $c = re^{i\theta}$

$$A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

y $p = (x - c)(x - \bar{c})$. Luego p es irreducible sobre R. Como p es divisible por el polinomio minimal de T, se sigue que p es el polinomio minimal.

Este ejemplo sugiere el siguiente recíproco.

Teorema 18. *Sean T un operador normal sobre un espacio producto interno real de dimensión finita V y p su polinomio minimal. Supóngase que*

$$p = (x - a)^2 + b^2$$

donde a y b son reales y $b \neq 0$. Entonces existe un entero $s > 0$ tal que p^s es el polinomio propio de T y existen subespacios $V_1, \ldots, V_s$ de V tales que

(i) *V_j es ortogonal a V_i cuando $i \neq j$;*
(ii) *$V = V_1 \oplus \cdots \oplus V_s$;*
(iii) *cada V_j tiene una base ortonormal $\{\alpha_j, \beta_j\}$ con la propiedad de que*

$$T\alpha_j = a\alpha_j + b\beta_j$$
$$T\beta_j = -b\alpha_j + a\beta_j.$$

En otras palabras, si $r = \sqrt{a^2 + b^2}$ y θ se elige de modo que $a = r \cos \theta$ y $b = r \operatorname{sen} \theta$, entonces V es una suma directa ortonormal de espacios bidimensionales V_j en cada uno de los cuales T actúa como «r veces la rotación de ángulo θ».

La demostración del Teorema 18 estará basada en el siguiente resultado.

Lema. *Sean V un espacio producto interno real y S un operador normal sobre V tal que $S^2 + I = 0$. Sea α cualquier vector de V y $\beta = S\alpha$. Entonces*

$$S^*\alpha = -\beta \qquad \text{(9-17)}$$
$$S^*\beta = \alpha$$

$(\alpha|\beta) = 0$, *y* $\|\alpha\| = \|\beta\|$.

Demostración. Tenemos que $S\alpha = \beta$ y $S\beta = S^2\alpha = -\alpha$. Por tanto,

$$0 = \|S\alpha - \beta\|^2 + \|S\beta + \alpha\|^2 = \|S\alpha\|^2 - 2(S\alpha|\beta) + \|\beta\|^2 + \|S\beta\|^2 + 2(S\beta|\alpha) + \|\alpha\|^2.$$

Como S es normal, se sigue que

$$0 = \|S^*\alpha\|^2 - 2(S^*\beta|\alpha) + \|\beta\|^2 + \|S^*\beta\|^2 + 2(S^*\alpha|\beta) + \|\alpha\|^2 = \|S^*\alpha + \beta\|^2 + \|S^*\beta - \alpha\|^2.$$

Esto implica (9-17); luego

$$(\alpha|\beta) = (S^*\beta|\beta) = (\beta|S\beta) = (\beta|-\alpha) = -(\alpha|\beta)$$

y $(\alpha|\beta) = 0$. En forma análoga

$$\|\alpha\|^2 = (S^*\beta|\alpha) = (\beta|S\alpha) = \|\beta\|^2. \quad ∎$$

Demostración del Teorema 18. Sea $V_1, \ldots, V_s$ una colección maximal de subespacios bidimensionales que satisfacen (i) y (ii) y las condiciones adicionales

$$T^*\alpha_j = a\alpha_j - b\beta_j, \qquad 1 \leq j \leq s. \qquad \text{(9-18)}$$
$$T^*\beta_j = b\alpha_j + a\beta_j$$

Sea $W = V_1 + \cdots + V_s$. Entonces W es la suma directa ortogonal de los

$V_1, \ldots, V_s$. Veremos que $W = V$. Supóngase que éste no sea el caso. Entonces $W^\perp \neq \{0\}$. Más aún, como (iii) y (9-18) implican que W es invariante por T y T^*, se sigue que $W^\perp$ es invariante por T^* y $T = T^{**}$. Sea $S = b^{-1}(T - aI)$. Entonces $S^* = b^{-1}(T^* - aI)$, $S^*S = SS^*$ y $W^\perp$ es invariante por S y S^*. Como $(T - aI)^2 + b^2I = 0$, se sigue que $S^2 + I = 0$. Sea α cualquier vector de norma 1 en $W^\perp$ y sea $\beta = S\alpha$. Entonces β está en $W^\perp$ y $S\beta = -\alpha$. Como $T = aI = bS$, esto implica que

$$\begin{aligned} T\alpha &= a\alpha + b\beta \\ T\beta &= -b\alpha + a\beta. \end{aligned}$$

Por el teorema, $S^*\alpha = -\beta$, $S^*\beta = \alpha$. $(\alpha|\beta) = 0$ y $\|\beta\| = 1$. Ya que $T^* = aI + bS^*$, se sigue que

$$\begin{aligned} T^*\alpha &= a\alpha - b\beta \\ T^*\beta &= b\alpha + a\beta. \end{aligned}$$

Pero esto contradice el hecho de que $V_1, \ldots, V_s$ es una colección maximal de subespacios que satisfacen (i), (iii) y (9-18). Por tanto, $W = V$, y como

$$\det \begin{bmatrix} x - a & b \\ -b & x - a \end{bmatrix} = (x - a)^2 + b^2$$

se sigue de (i), (ii) y (iii) que

$$\det (xI - T) = [(x - a)^2 + b^2]^s. \quad \blacksquare$$

Corolario. *En las condiciones del teorema, T es inversible y*

$$T^* = (a^2 + b^2)T^{-1}.$$

Demostración. Como

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} a & b \\ -b & a \end{bmatrix} = \begin{bmatrix} a^2 + b^2 & 0 \\ 0 & a^2 + b^2 \end{bmatrix}$$

se sigue de (iii) y (9-18) que $TT^* = (a^2 + b^2)I$. Luego T es inversible y $T^* = (a^2 + b^2)T^{-1}$. $\blacksquare$

Teorema 19. *Sea T un operador normal sobre un espacio producto interno de dimensión finita V. Entonces cualquier operador lineal que conmuta con T también conmuta con T^*. Además, todo subespacio invariante por T es también invariante por T^*.*

Demostración. Supóngase que U es un operador lineal sobre V que conmuta con T. Sea E_j la proyección ortogonal de V sobre el componente primo W_j $(1 \leq j \leq k)$ de V según T. Entonces E_j es un polinomio en T y, por tanto, conmuta con U. Así que

$$E_jUE_j = UE_j^2 = UE_j.$$

con lo que $U(W_j)$ es un subconjunto de W_j. Sean T_j y U_j las restricciones de T y U a W_j. Supóngase que I_j es el operador identidad sobre W_j. Entonces U_j conmuta con T_j, y si $T_j = c_jI_j$ es claro que U_j también conmuta con $T_j^* = \bar{c}_jI_j$.

Por otro lado, si T_j no es un múltiplo escalar de I_j, entonces T_j es inversible y existen números reales a_j y b_j tales que

$$T_j^* = (a_j^2 + b_j^2)T_j^{-1}$$

Como $U_jT_j = T_jU_j$ se sigue que $T_j^{-1}U_j = U_jT_j^{-1}$. Por tanto, U_j conmuta con T_j^* en ambos casos. Ahora T^* también conmuta con E_j y, por tanto, W_j es invariante por T^*. Además, para todo α y β de W_j

$$(T_j\alpha|\beta) = (T\alpha|\beta) = (\alpha|T^*\beta) = (\alpha|T_j^*\beta).$$

Como $T^*(W_j)$ está contenido en W_j, esto implica que T_j^* es la restricción de T^* a W_j. Así que

$$UT^*\alpha_j = T^*U\alpha_j$$

para todo α_j de W_j. Como V es la suma de $W_1, \ldots, W_k$, se sigue que

$$UT^*\alpha = T^*U\alpha$$

para todo α de V y, por tanto, que U conmuta con T^*.

Supóngase ahora que W es un subespacio de V invariante por T y sea $Z_j = W \cap W_j$. Por el corolario del Teorema 17, $W = \sum_j Z_j$. Así que es suficiente demostrar que todo Z_j es invariante por T_j^*. Esto es obvio si $T_j = c_jI$. Cuando éste no es el caso, T_j es inversible y aplica Z_j en, y en consecuencia sobre, Z_j. Así, $T_j^{-1}(Z_j) = Z_j$, y como

$$T_j^* = (a_j^2 + b_j^2)T_j^{-1},$$

se concluye que $T^*(Z_j)$ está contenido en Z_j, para todo j. ∎

Supóngase que T es un operador normal en un espacio producto interno de dimensión finita V. Sea W un subespacio invariante por T. Entonces el corolario anterior muestra que W es invariante por T^*. De esto se sigue que $W^\perp$ es invariante por $T^{**} = T$ (y, en consecuencia, también invariante por T^*). Usando este hecho se puede fácilmente demostrar la siguiente versión más reforzada del teorema de descomposición cíclica, dado en el Capítulo 7.

Teorema 20. *Sea T un operador lineal normal en un espacio producto interno de dimensión finita V ($\dim V \geq 1$). Entonces existen r vectores no nulos $\alpha_1, \ldots, \alpha_r$ en V con los respectivos T-anuladores $e_1, \ldots, e_r$ tales que*

(i) $V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T)$;

(ii) *si* $1 \leq k \leq r - 1$, *entonces* e_{k+1} *divide a* e_k;

(iii) $Z(\alpha_j; T)$ *es ortogonal a* $Z(\alpha_k; T)$ *si* $j \neq k$. *Además, el entero r y los anuladores $e_1, \ldots, e_r$ están unívocamente determinados por las condiciones* (i) *y* (ii) *y por el hecho de que ningún α_k es* 0.

Corolario. *Si A es una matriz normal de elementos reales (complejos), entonces existe una matriz real ortogonal (unitaria) P tal que $P^{-1}AP$ está en forma canónica racional.*

Se sigue que dos matrices A y B son equivalentes unitariamente si, y solo si, tienen la misma forma racional; A y B son equivalentes ortogonalmente si tienen elementos reales y la misma forma racional.

Por otro lado, hay un criterio más simple que la equivalencia unitaria de las matrices normales y los operadores normales.

Definiciones. *Sean V y V' espacios producto interno sobre el mismo cuerpo. Una transformación lineal*

$$U : V \rightarrow V'$$

se dice una **transformación unitaria** *si aplica V sobre V' y preserva los productos internos. Si T es un operador lineal sobre V y T' un operador lineal sobre V', entonces T es* **unitariamente equivalente** *a T', si existe una transformación unitaria U de V sobre V' tal que*

$$UTU^{-1} = T'.$$

Lema. *Sean V y V' espacios producto interno de dimensión finita sobre el mismo cuerpo. Supóngase que T es un operador lineal sobre V y que T' es un operador lineal sobre V'. Entonces T es unitariamente equivalente a T' si, y solo si, existen una base ortonormal $\mathcal{B}$ de V y una base ortonormal $\mathcal{B}'$ de V' tal que*

$$[T]_{\mathcal{B}} = [T']_{\mathcal{B}'}.$$

Demostración. Supóngase que hay una transformación unitaria U de V sobre V' tal que $UTU^{-1} = T'$. Sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ cualquier base (ordenada) ortonormal de V. Sea $\alpha'_j = U\alpha_j$ $(1 \leq j \leq n)$. Entonces $\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ es una base ortonormal de V', y haciendo

$$T\alpha_j = \sum_{k=1}^{n} A_{kj}\alpha_k$$

se ve que

$$\begin{aligned} T'\alpha'_j &= UT\alpha_j \\ &= \sum_k A_{kj}U\alpha_k \\ &= \sum_k A_{kj}\alpha'_k \end{aligned}$$

Luego $[T]_{\mathcal{B}} = A = [T']_{\mathcal{B}'}$.

Recíprocamente, supóngase que existen una base ortonormal $\mathcal{B}$ de V y una base ortonormal $\mathcal{B}'$ de V' tales que

$$[T]_{\mathcal{B}} = [T']_{\mathcal{B}'}$$

y sea $A = [T]_{\mathcal{B}}$. Supóngase que $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ y que $\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$. Sea U una transformación lineal de V en V' tal que $U\alpha_j = \alpha'_j$ $(1 \leq j \leq n)$. Entonces U es una transformación unitaria de V sobre V', y

$$\begin{aligned} UTU^{-1}\alpha'_j &= UT\alpha_j \\ &= U\sum_k A_{kj}\alpha_k \\ &= \sum_k A_{kj}\alpha'_k. \end{aligned}$$

Por tanto, $UTU^{-1}\alpha'_j = T'\alpha'_j$ $(1 \leq j \leq n)$, y esto implica que $UTU^{-1} = T'$. ▌

Se sigue inmediatamente de este lema que los operadores unitariamente equivalentes en espacios de dimensión finita tienen el mismo polinomio propio. Para operadores normales el recíproco es válido.

Teorema 21. *Sean V y V' espacios producto interno de dimensión finita sobre el mismo cuerpo. Supóngase que T es un operador normal sobre V y que T' es un operador normal sobre V'. Entonces T es unitariamente equivalente a T' si, y solo si, T y T' tienen el mismo polinomio propio.*

Demostración. Supóngase que T y T' tienen el mismo polinomio propio f. Sean W_j $(1 \leq j \leq k)$ los componentes primos de V según T y T_j la restricción de T a W_j. Supóngase que I_j sea el operador identidad sobre W_j. Entonces

$$f = \prod_{j=1}^{k} \det (xI_j - T_j).$$

Sea p_j el polinomio minimal de T_j. Si $p_j = x - c_j$ es evidente que

$$\det (xI_j - T_j) = (x - c_j)^{s_j}$$

donde s_j es la dimensión de W_j. Por otro lado, si $p_j = (x - a_j)^2 + b_j^2$ con a_j, b_j reales y $b_j \neq 0$, entonces se sigue del Teorema 18 que

$$\det (xI_j - T_j) = p_j^{s_j}$$

donde en este caso $2s_j$ es la dimensión de W_j. Por consiguiente, $f = \prod_j p_j^{s_j}$. Se puede ahora calcular también f por el mismo método usando los componentes primos de V' según T'. Como $p_1, \ldots, p_k$ son primos distintos, se sigue de la unicidad de la factorización prima de f que hay exactamente k componentes primos W'_j $(1 \leq j \leq k)$ de V' según T' y que éstos pueden ser rotulados con índices de tal modo que p_j sea el polinomio minimal de la restricción T'_j de T' a W'_j. Si $p_j = x - c_j$, entonces $T_j = c_j I_j$ y $T'_j = c_j I'_j$, donde I'_j es el operador identidad sobre W'_j. En este caso es evidente que T_j es unitariamente equivalente a T'_j. Si $p_j = (x - a_j)^2 + b_j^2$, como anteriormente, entonces por el lema y el Teorema 20, tenemos nuevamente que T_j es unitariamente equivalente a T'_j. Así, pues, para todo j existen bases ortonormales $\mathcal{B}_j$ y $\mathcal{B}'_j$ de W_j y W'_j, respectivamente, tales que

$$[T_j]_{\mathcal{B}_j} = [T'_j]_{\mathcal{B}'_j}.$$

Sea ahora U la transformación lineal de V en V' que aplica cada $\mathcal{B}_j$ sobre $\mathcal{B}'_j$. Entonces U es una transformación unitaria de V sobre V' tal que $UTU^{-1} = T'$. ▌

10. Formas bilineales

10.1. Formas bilineales

En este capítulo se tratará de las formas bilineales sobre espacios vectoriales de dimensión finita. El lector observará probablemente una analogía entre cierto material y el estudio de los determinantes del Capítulo 5 y de los productos internos y de las formas en los Capítulos 8 y 9. La relación entre formas bilineales y productos internos es particularmente estrecha; sin embargo, este capítulo no presupone ningún material de los Capítulos 8 o 9. Al lector que no esté familiarizado con los productos internos le será provechoso leer la primera parte del Capítulo 8 al tiempo que se adentra en el estudio de las formas bilineales.

La primera sección trata del espacio de las formas bilineales sobre un espacio vectorial de dimensión n. Se introduce la matriz de una forma bilineal en una base ordenada y se establece el isomorfismo entre el espacio de las formas y el espacio de las matrices $n \times n$. Se define el rango de una forma bilineal y se introducen las formas bilineales no degeneradas. La segunda sección estudia las formas bilineales simétricas y su diagonalización. La tercera trata las formas bilineales antisimétricas. La cuarta estudia el grupo que preserva una forma bilineal no degenerada, con especial atención a los grupos ortogonales, a los grupos seudoortogonales y a un grupo seudoortogonal particular —el grupo de Lorentz.

Definición. *Sea V un espacio vectorial sobre el cuerpo F. Una* **forma bilineal** *sobre V es una función f que asigna a cada par ordenado de vectores α, β de V un escalar $f(\alpha, \beta)$ de F, y que satisface*

$$(10\text{-}1) \qquad \begin{aligned} f(c\alpha_1 + \alpha_2, \beta) &= cf(\alpha_1, \beta) + f(\alpha_2, \beta) \\ f(\alpha, c\beta_1 + \beta_2) &= cf(\alpha, \beta_1) + f(\alpha, \beta_2). \end{aligned}$$

Si $V \times V$ es el conjunto de todos los pares ordenados de vectores de V, esta definición puede expresarse como sigue: una forma bilineal sobre V es una función f de $V \times V$ en F que es lineal como función de uno de sus argumentos cuando el otro se deja fijo. La función cero de $V \times V$ en T es obviamente una forma bilineal. También es cierto que cualquier combinación lineal de formas bilineales sobre V es una forma bilineal. Para demostrarlo es suficiente considerar combinaciones lineales del tipo $cf + g$, donde f y g son formas bilineales sobre V. La demostración de que $cf + g$ satisface (10-1) es similar a muchas otras que se han dado, y por tal razón se omitirá. Todo esto puede resumirse diciendo que el conjunto de todas las formas bilineales sobre V es un subespacio del espacio de todas las funciones de $V \times V$ en F (Ejemplo 3, Capítulo 2). Se representará el espacio de las formas bilineales sobre V por $L(V, V, F)$.

Ejemplo 1. Sea V un espacio vectorial sobre el cuerpo F y sean L_1 y L_2 formas bilineales en V. Se define f por

$$f(\alpha, \beta) = L_1(\alpha)L_2(\beta).$$

Si se fija β y se considera a f como función de α, entonces se tiene simplemente un múltiplo escalar del funcional lineal L_1. Con α fijo, f es un múltiplo escalar de L_2. Así es claro que f es una forma bilineal en V.

Ejemplo 2. Sean m y n enteros positivos y F un cuerpo. Sea V el espacio vectorial de las matrices $m \times n$ sobre F. Sea A una matriz $m \times m$ dada sobre F. Se define

$$f_A(X, Y) = \operatorname{tr}(X^tAY).$$

Entonces f_A es una forma bilineal sobre V. En efecto, si X, Y y Z son matrices $m \times n$ sobre F

$$\begin{aligned} f_A(cX + Z, Y) &= \operatorname{tr}[(cX + Z)^tAY] \\ &= \operatorname{tr}(cXtAY) + \operatorname{tr}(Z^tAY) \\ &= cf_A(X, Y) + f_A(Z, Y). \end{aligned}$$

Desde luego, se ha hecho uso de que la operación transpuesta y la función traza son lineales. Es incluso más fácil demostrar que f_A es lineal como función de su segundo argumento. En el caso especial $n = 1$, la matriz X^tAY es 1×1, es decir, un escalar, y la forma bilineal es solamente

$$\begin{aligned} f_A(X, Y) &= X^tAY \\ &= \sum_i \sum_j A_{ij}x_iy_j. \end{aligned}$$

Mostraremos ahora que toda forma bilineal sobre el espacio de las matrices $m \times 1$ es de este tipo; es decir, es f_A para cierta matriz $m \times m$, A.

Ejemplo 3. Sea F un cuerpo. Queremos hallar todas las formas bilineales sobre el espacio F^2. Supóngase que f sea una forma bilineal de este tipo. Si $\alpha = (x_1, x_2)$ y $\beta = (y_1, y_2)$ son vectores de F^2, entonces

$$\begin{aligned} f(\alpha, \beta) &= f(x_1\epsilon_1 + x_2\epsilon_2, \beta) \\ &= x_1 f(\epsilon_1, \beta) + x_2 f(\epsilon_2, \beta) \\ &= x_1 f(\epsilon_1, y_1\epsilon_1 + y_2\epsilon_2) + x_2 f(\epsilon_2, y_1\epsilon_1 + y_2\epsilon_2) \\ &= x_1y_1 f(\epsilon_1, \epsilon_1) + x_1y_2 f(\epsilon_1, \epsilon_2) + x_2y_1 f(\epsilon_2, \epsilon_1) + x_2y_2 f(\epsilon_2, \epsilon_2). \end{aligned}$$

Así f está completamente determinada por los cuatro escalares $A_{ij} = f(\epsilon_i, \epsilon_j)$ por

$$\begin{aligned} f(\alpha, \beta) &= A_{11}x_1y_1 + A_{12}x_1y_2 + A_{21}x_2y_1 + A_{22}x_2y_2 \\ &= \sum_{i,j} A_{ij}x_iy_j. \end{aligned}$$

Si X e Y son las matrices de coordenadas de α y β y si A es la matriz 2×2 con los elementos $A(i, j) = A_{ij} = f(\epsilon_i, \epsilon_j)$, entonces

$$f(\alpha, \beta) = X^tAY. \tag{10-2}$$

Se observó en el Ejemplo 2 que si A es cualquier matriz 2×2 sobre F, entonces (10-2) define una forma bilineal sobre F^2. Se ve que las formas bilineales sobre F^2 son precisamente las que se obtienen de una matriz 2×2 como en (10-2).

El análisis del Ejemplo 3 puede generalizarse para describir todas las formas bilineales sobre un espacio vectorial de dimensión finita. Sea V un espacio vectorial de dimensión finita sobre el cuerpo F y sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ordenada de V. Supóngase que f es una forma bilineal en V. Si

$$\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n \quad \text{y} \quad \beta = y_1\alpha_1 + \cdots + y_n\alpha_n$$

son vectores en V, entonces

$$\begin{aligned} f(\alpha, \beta) &= f\left(\sum_i x_i\alpha_i, \beta\right) \\ &= \sum_i x_i f(\alpha_i, \beta) \\ &= \sum_i x_i f\left(\alpha_i, \sum_j y_j\alpha_j\right) \\ &= \sum_i \sum_j x_iy_j f(\alpha_i, \alpha_j). \end{aligned}$$

Si se hace $A_{ij} = f(\alpha_i, \alpha_j)$, entonces

$$\begin{aligned} f(\alpha, \beta) &= \sum_i \sum_j A_{ij}x_iy_j \\ &= X^tAY \end{aligned}$$

donde X e Y son las matrices de coordenadas de α y β en la base ordenada $\mathcal{B}$. Así toda forma bilineal en V es del tipo

$$f(\alpha, \beta) = [\alpha]^t_{\mathcal{B}} A [\beta]_{\mathcal{B}} \tag{10-3}$$

para alguna matriz $n \times n$ sobre F. Recíprocamente, si se tiene cualquier matriz $n \times n$, A, es fácil ver que (10-3) define una forma bilineal f sobre V, tal que $A_{ij} = f(\alpha_i, \alpha_j)$.

Definición. *Sea V un espacio vectorial de dimensión finita y sea $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ una base ordenada de V. Si f es una forma bilineal sobre V, la* **matriz de f en la base ordenada** *$\mathcal{B}$ es la matriz $n \times n$, A, con elementos $A_{ij} = f(\alpha_i, \alpha_j)$. A veces se representará esta matriz por $[f]_{\mathcal{B}}$.*

Teorema 1. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo F. Para cada base ordenada $\mathcal{B}$ de V, la función que asocia a cada forma bilineal sobre V su matriz en la base ordenada $\mathcal{B}$ es un isomorfismo del espacio $L(V, V, F)$ sobre el espacio de las matrices $n \times n$ sobre el cuerpo F.*

Demostración. Se observó anteriormente que $f \to [f]_{\mathcal{B}}$ es una correspondencia biunívoca entre el conjunto de las formas bilineales sobre V y el conjunto de todas las matrices $n \times n$ sobre F. Que es una transformación lineal es fácil de ver, ya que

$$(cf + g)(\alpha_i, \alpha_j) = cf(\alpha_i, \alpha_j) + g(\alpha_i, \alpha_j)$$

para todo i y j. Esto simplemente dice que

$$[cf + g]_{\mathcal{B}} = c[f]_{\mathcal{B}} + [g]_{\mathcal{B}}. \quad \blacksquare$$

Corolario. *Si $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ es una base ordenada de V y $\mathcal{B}^* = \{L_1, \ldots, L_n\}$ es la base dual de V^*, entonces las n^2 formas bilineales*

$$f_{ij}(\alpha, \beta) = L_i(\alpha)L_j(\beta), \qquad 1 \leq i \leq n,\ 1 \leq j \leq n$$

forman una base del espacio $L(V, V, F)$. En particular, la dimensión de $L(V, V, F)$ es n^2.

Demostración. La base dual $\{L_1, \ldots, L_n\}$ está esencialmente definida por el hecho de que $L_i(\alpha)$ es la i-ésima coordenada de α en la base ordenada $\mathcal{B}$ (para todo α de V). Ahora bien, las funciones f_{ij} definidas por

$$f_{ij}(\alpha, \beta) = L_i(\alpha)L_j(\beta)$$

son formas bilineales del tipo considerado en el Ejemplo 1. Si

$$\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n \quad \text{y} \quad \beta = y_1\alpha_1 + \cdots + y_n\alpha_n,$$

entonces

$$f_{ij}(\alpha, \beta) = x_iy_j.$$

Sea f cualquier forma bilineal sobre V y sea A la matriz de f en la base ordenada $\mathcal{B}$. Entonces

$$f(\alpha, \beta) = \sum_{i,j} A_{ij}x_iy_j$$

que simplemente dice que

$$f = \sum_{i,j} A_{ij}f_{ij}.$$

Está ahora claro que las n^2 formas f_{ij} constituyen una base de $L(V, V, F)$. ▌

La demostración anterior se puede expresar de otra manera como sigue. La forma bilineal f_{ij} tiene por matriz en la base ordenada $\mathcal{B}$ la matriz «unitaria» $E^{i,j}$, cuyo único elemento no nulo es un 1 en la fila i y la columna j. Como estas matrices unitarias constituyen una base del espacio de las matrices $n \times n$, las formas f_{ij} constituyen una base del espacio de las formas bilineales.

El concepto de matriz de una forma bilineal en una base ordenada es análogo al de matriz de un operador lineal en una base ordenada. Al igual que para los operadores lineales, estamos interesados en qué le sucede a la matriz representante de una forma bilineal cuando se pasa de una base ordenada a otra. Para ello, supóngase que $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ y $\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ sean dos bases ordenadas para V y que f sea una forma bilineal en V. ¿Cómo están relacionadas las matrices $[f]_{\mathcal{B}}$ y $[f]_{\mathcal{B}'}$? Sea P la matriz $n \times n$ (inversible) tal que

$$[\alpha]_{\mathcal{B}} = P[\alpha]_{\mathcal{B}'}$$

para todo α en V. En otras palabras, se define P por

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Para vectores α, β cualesquiera en V

$$\begin{aligned} f(\alpha, \beta) &= [\alpha]^t_{\mathcal{B}}[f]_{\mathcal{B}}[\beta]_{\mathcal{B}} \\ &= (P[\alpha]_{\mathcal{B}'})^t[f]_{\mathcal{B}}P[\beta]_{\mathcal{B}'} \\ &= [\alpha]^t_{\mathcal{B}'}(P^t[f]_{\mathcal{B}}P)[\beta]_{\mathcal{B}'}. \end{aligned}$$

Por la definición y unicidad de la matriz representante de f en la base ordenada $\mathcal{B}'$, debemos tener que

$$(10\text{-}4) \qquad [f]_{\mathcal{B}'} = P^t[f]_{\mathcal{B}}P.$$

Ejemplo 4. Sea V el espacio vectorial R^2. Sea f la forma bilineal definida por $\alpha = (x_1, x_2)$ y $\beta = (y_1, y_2)$ por

$$f(\alpha, \beta) = x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2.$$

Ahora

$$f(\alpha, \beta) = [x_1, x_2]\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

y así la matriz de f en la base ordenada canónica $\mathcal{B} = \{\epsilon_1, \epsilon_2\}$ es

$$[f]_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Sea $\mathcal{B}' = \{\epsilon'_1, \epsilon'_2\}$ la base ordenada definida por $\epsilon'_1 = (1, -1)$, $\epsilon'_2 = (1, 1)$. En este caso, la matriz P que cambia las coordenadas de $\mathcal{B}'$ a $\mathcal{B}$ es

$$P = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Así

$$\begin{aligned}[f]_{\mathfrak{B}'} &= P^t[f]_{\mathfrak{B}}P\\ &= \begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}\\ &= \begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}0 & 2\\ 0 & 2\end{bmatrix}\\ &= \begin{bmatrix}0 & 0\\ 0 & 4\end{bmatrix}.\end{aligned}$$

Esto quiere decir que si se expresan los vectores α y β por medio de sus coordenadas en la base $\mathfrak{B}'$, por ejemplo,

$$\alpha = x_1'\epsilon_1' + x_2'\epsilon_2', \qquad \beta = y_1'\epsilon_1' + y_2'\epsilon_2'$$

entonces

$$f(\alpha, \beta) = 4x_2'y_2'.$$

Una consecuencia de la fórmula (10-4) para el cambio de base es la siguiente: si A y B son matrices $n \times n$ que representan la misma forma bilineal sobre V en bases ordenadas (posiblemente) diferentes, entonces A y B tienen el mismo rango. En efecto, si P es una matriz $n \times n$ inversible y $B = P^tAP$, es evidente que A y B tienen el mismo rango. Esto hace posible definir el rango de una forma bilineal en V como el rango de cualquier matriz que represente la forma en una base ordenada para V.

Es deseable dar una definición más intrínseca del rango de una forma bilineal. Esto se puede hacer del siguiente modo. Supóngase que f es una forma bilineal sobre el espacio vectorial V. Si se fija un vector α de V, entonces $f(\alpha, \beta)$ es lineal como función de β. De este modo cada α fijo determina un funcional lineal en V_j; se representa esta función lineal por $L_j(\alpha)$. Repitiendo, si α es un vector en V, entonces $L_j(\alpha)$ es el funcional lineal en V cuyo valor para cualquier vector β es $f(\alpha, \beta)$. Esto da una transformación $\alpha \rightarrow L_j(\alpha)$ de V en el espacio dual V^*. Como

$$f(c\alpha_1 + \alpha_2, \beta) = cf(\alpha_1, \beta) + f(\alpha_2, \beta)$$

vemos que

$$L_f(c\alpha_1 + \alpha_2) = cL_f(\alpha_1) + L_f(\alpha_2)$$

esto es, L_j es una transformación lineal de V en V^*.

De modo semejante, f determina una transformación lineal R_f de V en V^*. Para cada β fijo en V, $f(\alpha, \beta)$ es lineal como función de α. Se define $R_f(\beta)$ como el funcional lineal sobre V cuyo valor para el vector α es $f(\alpha, \beta)$.

Teorema 2. *Sea f una forma bilineal sobre el espacio vectorial de dimensión finita V. Sean L_f y R_f las transformaciones lineales de V en V^* definidas por $(L_f\alpha)(\beta) = f(\alpha, \beta) = (R_f\beta)(\alpha)$. Entonces, rango (L_f) = rango (R_f).*

Demostración. Se puede dar una demostración de este teorema que no dependa de las coordenadas. Tal demostración es parecida a la demostración

de que el rango de fila de una matriz es igual al rango de columna (Sección 3.7). Así que se dará aquí una demostración que consiste en elegir un sistema de coordenadas (base) y entonces usar el teorema de «rango de fila igual a rango de columna».

Para demostrar que rango (L_f) = rango (R_f) es suficiente demostrar que L_f y R_f tienen la misma nulidad. Sea $\mathcal{B}$ una base ordenada para V y sea $A = [f]_{\mathcal{B}}$. Si α y β son vectores en V, con matrices de coordenadas X e Y en la base ordenada $\mathcal{B}$, entonces $f(\alpha, \beta) = X^tAY$. Ahora, $R_f(\beta) = 0$ quiere decir que $f(\alpha, \beta) = 0$ para todo α en V, es decir, que $X^tAY = 0$ para toda matriz $n \times 1$, X. Esta última condición dice que $AY = 0$. La nulidad de R_f es, por tanto, igual a la dimensión del espacio solución de $AY = 0$.

Similarmente, $L_f(\alpha) = 0$ si, y solo si, $X^tAY = 0$ para toda matriz $n \times 1$, Y. Así α está en el espacio nulo de L_f si, y solo si, $X^tA = 0$ es decir, $A^tX = 0$. La nulidad de L_f es, por tanto, igual a la dimensión del espacio solución de $A^tX = 0$. Como las matrices A y A^t tienen el mismo rango columna, vemos que

$$\text{nulidad } (L_f) = \text{nulidad } (R_f). \quad \blacksquare$$

Definición. *Si f es una forma bilineal sobre un espacio de dimensión finita V, el* **rango** *de f es el entero r = rango (L_f) = rango (R_f).*

Corolario 1. *El rango de una forma bilineal es igual al rango de la matriz de la forma en cualquier base ordenada.*

Corolario 2. *Si f es una forma bilineal sobre el espacio vectorial de dimensión n, lo siguiente es equivalente:*

(a) *Rango $(f) = n$.*
(b) *Para todo α no nulo de V, existe un β en V tal que $f(\alpha, \beta) \neq 0$.*
(c) *Para todo β no nulo de V, existe un α en V tal que $f(\alpha, \beta) \neq 0$.*

Demostración. La afirmación (b) no dice más que el espacio nulo de L_f es el subespacio cero. La afirmación (c) dice que el espacio nulo de R_f es el subespacio cero. Las transformaciones lineales L_f y R_f tienen nulidad 0 si, y solo si, tienen rango n, es decir, si, y solo si, rango $(f) = n$. $\blacksquare$

Definición. *Una forma bilineal f sobre un espacio vectorial V se llama* **no degenerada** (*o* **no singular**) *si satisface las condiciones* (b) *y* (c) *del Corolario* 2.

Si V es de dimensión finita, entonces f es no degenerada siempre que f cumpla una cualquiera de las tres condiciones del Corolario 2. En particular, f es no degenerada (no singular) si, y solo si, su matriz en alguna (toda) base ordenada para V es una matriz no singular.

Ejemplo 5. Sea $V = R^n$ y sea f la forma bilineal definida para $\alpha = (x_1, \ldots, x_n)$ y $\beta = (y_1, \ldots, y_n)$ por

$$f(\alpha, \beta) = x_1y_1 + \cdots + x_ny_n.$$

Entonces f es una forma bilineal no degenerada en R^n. La matriz de f en la base ordenada canónica es la matriz unidad $n \times n$

$$f(X, Y) = X^tY.$$

Este f generalmente se llama producto escalar. El lector estará probablemente familiarizado con esta forma bilineal, al menos para el caso $n = 3$. Geométricamente el número $f(\alpha, \beta)$ es el producto de la longitud de α, la longitud de β y el coseno del ángulo entre α y β. En particular $f(\alpha, \beta) = 0$ si, y solo si, los vectores α y β son ortogonales (perpendiculares).

Ejercicios

1. ¿Cuáles de las siguientes funciones f definidas para vectores $\alpha = (x_1, x_2)$ y $\beta = (y_1, y_2)$ de R^2 son formas bilineales?

(a) $f(\alpha, \beta) = 1.$
(b) $f(\alpha, \beta) = (x_1 - y_1)^2 + x_2y_2.$
(c) $f(\alpha, \beta) = (x_1 + y_1)^2 - (x_1 - y_1)^2.$
(d) $f(\alpha, \beta) = x_1y_2 - x_2y_1.$

2. Sea f la forma bilineal sobre R^2 definida por

$$f((x_1, y_1), (x_2, y_2)) = x_1y_1 + x_2y_2.$$

Hallar la matriz de f en cada una de las siguientes bases

$$\{(1, 0), (0, 1)\}, \qquad \{(1, -1), (1, 1)\}, \qquad \{(1, 2), (3, 4)\}.$$

3. Sea V el espacio de todas las matrices 2×3 sobre R y sea f la forma bilineal sobre V definida por $f(X, Y) = \text{traza}\,(X^tAY)$, donde

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Hallar la matriz de f en la base ordenada

$$\{E^{11}, E^{12}, E^{13}, E^{21}, E^{22}, E^{23}\}$$

donde E^{ij} es la matriz cuyo único elemento no nulo es un 1 en la fila i y columna j.

4. Describir explícitamente todas las formas bilineales f sobre R^3 con la propiedad de que $f(\alpha, \beta) = f(\beta, \alpha)$ para todo α, β.

5. Describir la forma bilineal sobre R^2 que satisface $f(\alpha, \beta) = -f(\beta, \alpha)$ para todo α, β.

6. Sea n un entero positivo y sea V el espacio de todas las matrices $n \times n$ sobre el cuerpo de los números complejos. Demostrar que

$$f(A, B) = n\,\text{tr}\,(AB) - \text{tr}\,(A)\,\text{tr}\,(B)$$

define una forma bilineal f sobre V. ¿Es verdad que $f(A, B) = f(B, A)$ para toda A, B?

7. Sea f la forma bilineal definida en el Ejercicio 6. Demostrar que f es degenerada (que no es no degenerada). Sea V_1 el subespacio de V que consiste en las matrices de traza 0, y sea f_1 la restricción de f a V_1. Mostrar que f_1 es no degenerada.

8. Sea f la forma bilineal definida en el Ejercicio 6 y sea V_2 el subespacio de V que consis-

te en todas las matrices A tal que traza $(A) = 0$ y $A^* = -A$ (A^* es la transpuesta conjugada de A). Desígnese por f_2 la restricción de f a V_2. Mostrar que f_2 es negativamente definida; es decir, que $f_2(A, A) < 0$ para cada A no nulo en V_2.

9. Sea f la forma bilineal definida en el Ejercicio 6. Sea W el conjunto de todas las matrices A en V tales que $f(A, B) = 0$ para todo B. Demostrar que W es un subespacio de V. Describir W explícitamente y hallar su dimensión.

10. Sea f cualquier forma bilineal sobre un espacio vectorial de dimensión finita V. Sea W el subespacio de todos los β tales que $f(\alpha, \beta) = 0$ para todo α. Mostrar que

$$\text{rango } f = \dim V - \dim W.$$

Usar este resultado y el del Ejercicio 9 para calcular el rango de la forma bilineal definida en el Ejercicio 6.

11. Sea f una forma bilineal sobre un espacio vectorial de dimensión finita V. Supóngase que V_1 es un subespacio de V con la propiedad de que la restricción de f a V_1 es no degenerada. Demostrar que rango $f \geq \dim V_1$.

12. Sean f, g formas bilineales sobre un espacio vectorial de dimensión finita V. Supóngase que g es no singular. Demostrar que existen operadores lineales únicos T_1, T_2 sobre V tales que

$$f(\alpha, \beta) = g(T_1\alpha, \beta) = g(\alpha, T_2\beta)$$

para todo α, β.

13. Demostrar que el resultado dado en el Ejercicio 12 no necesita ser cierto si g es singular.

14. Sea f una forma bilineal sobre un espacio vectorial de dimensión finita V. Demostrar que f puede expresarse como producto de dos funcionales lineales (es decir, $f(\alpha, \beta) = L_1(\alpha)L_2(\beta)$ para L_1, L_2 en V^*) si, y solo si, f tiene rango 1.

10.2. Formas bilineales simétricas

El propósito principal de esta sección es responder a la siguiente pregunta: si f es una forma bilineal sobre el espacio vectorial de dimensión finita V, ¿cuándo existe una base ordenada $\mathcal{B}$ de V en la que f esté representada por una matriz diagonal? Se demostrará que esto es posible si, y solo si, f es una forma bilineal simétrica, es decir, $f(\alpha, \beta) = f(\beta, \alpha)$. El teorema se demostrará solo cuando el cuerpo de los escalares tiene característica cero, esto es, que si n es un entero positivo la suma $1 + 1 + \cdots + 1$ (n sumandos) en F no es 0.

Definición. *Sea f una forma bilineal sobre el espacio vectorial V. Se dice que f es* **simétrica** *si $f(\alpha, \beta) = f(\beta, \alpha)$ para todos los vectores α, β de V.*

Si V es de dimensión finita, la forma bilineal f es simétrica si, y solo si, su matriz A en cierta (o en toda) base ordenada es simétrica, $A^t = A$. Para ver esto se requiere saber cuándo la forma bilineal

$$f(X, Y) = X^tAY$$

es simétrica. Esto sucede si, y solo si, $X^tAY = Y^tAX$ para todas las matrices columnas X e Y. Como X^tAY es una matriz 1×1, se tiene que $X^tAY = Y^tA^tX$. Así, pues, f es simétrica si, y solo si, $Y^tA^tX = Y^tAX$ para todo X, Y. Evidentemente esto quiere decir que $A = A^t$. En particular, debe observarse que si existe una base ordenada para V en la que f está representada por una matriz diagonal, entonces f es simétrica, ya que cualquier matriz diagonal es una matriz simétrica.

Si f es una forma bilineal simétrica, la **forma cuadrática asociada con** f es la función q de V en F definida por

$$q(\alpha) = f(\alpha, \alpha).$$

Si F es un subconjunto de los números complejos, la forma bilineal simétrica está completamente determinada por su forma cuadrática asociada, de acuerdo con la **identidad de polarización**

$$f(\alpha, \beta) = \tfrac{1}{4}q(\alpha + \beta) - \tfrac{1}{4}q(\alpha - \beta). \tag{10-5}$$

El establecer (10-5) no es más que una rutina de cálculo que se omite. Si f es la forma bilineal del Ejemplo 5, el producto escalar, la forma cuadrática asociada es

$$q(x_1, \ldots, x_n) = x_1^2 + \cdots + x_n^2.$$

En otras palabras, $q(\alpha)$ es el cuadrado de la longitud de α. Para la forma bilineal $f_A(X, Y) = X^tAY$, la forma cuadrática asociada es

$$q_A(X) = X^tAX = \sum_{i,j} A_{ij}x_ix_j.$$

Una clase importante de formas bilineales simétricas son los productos internos en los espacios vectoriales reales, estudiados en el Capítulo 8. Si V es un espacio vectorial *real*, un **producto interno** sobre V es una forma bilineal simétrica f sobre V tal que

$$f(\alpha, \alpha) > 0 \quad \text{si} \quad \alpha \neq 0. \tag{10-6}$$

Una forma bilineal que cumple (10-6) se llama **positivamente definida.** Así, un producto interno sobre un espacio vectorial real es una forma bilineal simétrica y positivamente definida sobre ese espacio. Obsérvese que un producto interno es no degenerado. Dos vectores α, β se llaman **ortogonales** respecto al producto interno f si $f(\alpha, \beta) = 0$. La forma cuadrática $q(\alpha) = f(\alpha, \alpha)$ toma solo valores no negativos y $q(\alpha)$ se considera corrientemente como el cuadrado de la longitud de α. Desde luego, estos conceptos de longitud y ortogonalidad tienen su origen en el más importante ejemplo de un producto interno —el producto escalar del Ejemplo 5.

Si f es cualquier forma bilineal simétrica sobre un espacio vectorial V, es conveniente aplicar en parte la terminología a los productos internos a f. Es especialmente conveniente decir que α y β son ortogonales con respecto a f, si $f(\alpha, \beta) = 0$. No es aconsejable pensar en $f(\alpha, \alpha)$ como el cuadrado de la longitud de α; por ejemplo, si V es un espacio vectorial complejo, se puede tener $f(\alpha, \alpha) = \sqrt{-1}$, o en un espacio vectorial real, $f(\alpha, \alpha) = -2$.

Se vuelve ahora al teorema básico de esta sección. Al leer la demostración, le será de provecho al lector pensar en el caso especial en que V es un espacio vectorial real y f un producto interno sobre V.

Teorema 3. *Sea V un espacio vectorial de dimensión finita sobre un cuerpo de característica cero, y sea f una forma bilineal simétrica sobre V. Entonces existe una base de V en la que f está representada por una matriz diagonal.*

Demostración. Lo que se debe hallar es una base ordenada

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$$

tal que $f(\alpha_i, \alpha_j) = 0$ para $i \neq j$. Si $f = 0$ o $n = 1$, el teorema es evidentemente cierto. Por tanto, se puede suponer que $f \neq 0$ y $n > 1$. Si $f(\alpha, \alpha) = 0$ para todo α en V, la forma cuadrática asociada q es idénticamente 0 y la identidad de polarización (10-5) muestra que $f = 0$. Entonces existe un vector α en V tal que $f(\alpha, \alpha) = q(\alpha) \neq 0$. Sea W el subespacio unidimensional de V que es generado por α y sea $W^\perp$ el conjunto de todos los vectores β en V tales que $f(\alpha, \beta) = 0$. Afirmamos que $V = W \oplus W^\perp$. Ciertamente, los subespacios W y $W^\perp$ son independientes. Un vector típico de W es $c\alpha$, donde c es un escalar. Si también $c\alpha$ está en $W^\perp$, entonces $f(c\alpha, c\alpha) = c^2 f(\alpha, \alpha) = 0$. Pero $f(\alpha, \alpha) \neq 0$, luego $c = 0$. Así que cada vector de V es suma de un vector de W y de un vector de $W^\perp$. En efecto, sea γ cualquier vector en V, y se hace

$$\beta = \gamma - \frac{f(\gamma, \alpha)}{f(\alpha, \alpha)} \alpha.$$

Entonces

$$f(\alpha, \beta) = f(\alpha, \gamma) - \frac{f(\gamma, \alpha)}{f(\alpha, \alpha)} f(\alpha, \alpha)$$

y como f es simétrica, $f(\alpha, \beta) = 0$. Así β está en el subespacio $W^\perp$. La expresión

$$\gamma = \frac{f(\gamma, \alpha)}{f(\alpha, \alpha)} \alpha + \beta$$

muestra que $V = W + W^\perp$.

La restricción de f a $W^\perp$ es una forma bilineal simétrica. Como $W^\perp$ tiene dimensión $(n - 1)$, suponer por inducción que $W^\perp$ tiene una base $\{\alpha_2, \ldots, \alpha_n\}$ tal que

$$f(\alpha_i, \alpha_j) = 0, \qquad i \neq j \ (i \geq 2, j \geq 2).$$

Haciendo $\alpha_1 = \alpha$, se obtiene una base $\{\alpha_1, \ldots, \alpha_n\}$ de V tal que $f(\alpha_i, \alpha_j) = 0$ para $i \neq j$. ∎

Corolario. *Sea F un subcuerpo de los números complejos, y sea A una matriz simétrica $n \times n$ sobre F. Entonces existe una matriz $n \times n$ inversible P sobre F tal que P^tAP es diagonal.*

En caso de que F sea el cuerpo de los números reales, la matriz inversible P de este corolario puede elegirse de modo que sea una matriz *ortogonal*, es

decir, $P' = P^{-1}$. En otras palabras, si A es una matriz simétrica real $n \times n$, existe una matriz ortogonal real P tal que $P'AP$ es diagonal; sin embargo, esto no es del todo evidente en lo que se hizo anteriormente (véase Capítulo 8).

Teorema 4. *Sea V un espacio vectorial de dimensión finita sobre el cuerpo de los números complejos. Sea f una forma bilineal simétrica sobre V que tiene rango r. Entonces existe una base ordenada $\mathcal{B} = \{\beta_1, \ldots, \beta_n\}$ de V tal que*

(i) *la matriz de f en la base ordenada $\mathcal{B}$ es diagonal;*

(ii) $f(\beta_j, \beta_j) = \begin{cases} 1, & j = 1, \ldots, r \\ 0, & j > r. \end{cases}$

Demostración. Por el Teorema 3, existe una base ordenada $\{\alpha_1, \ldots, \alpha_n\}$ de V tal que

$$f(\alpha_i, \alpha_j) = 0 \quad \text{para} \quad i \neq j.$$

Como f tiene rango r, también lo tendrá su matriz en la base ordenada $\{\alpha_1, \ldots, \alpha_n\}$. Entonces debemos tener que $f(\alpha_j, \alpha_j) \neq 0$ para r valores de j precisamente. Reordenando los vectores α_j, se puede suponer que

$$f(\alpha_j, \alpha_j) \neq 0, \qquad j = 1, \ldots, r.$$

Se usa ahora el hecho de que el cuerpo de los escalares es el cuerpo de los números complejos. Si $\sqrt{f(\alpha_j, \alpha_j)}$ representa cualquier raíz cuadrada compleja de $f(\alpha_j, \alpha_j)$, y si se hace

$$\beta_j = \begin{cases} \dfrac{1}{\sqrt{f(\alpha_j, \alpha_j)}}\,\alpha_j, & j = 1, \ldots, r \\ \alpha_j, & j > r \end{cases}$$

la base $\{\beta_1, \ldots, \beta_n\}$ satisface las condiciones (i) y (ii). ∎

Naturalmente, el Teorema 4 es válido si el cuerpo de los escalares es cualquier subcuerpo de los números complejos en el que cada elemento tiene una raíz cuadrada. No es válido, cuando el cuerpo escalar es el cuerpo de los números reales. Sobre el cuerpo de los números reales tenemos el siguiente sustituto para el Teorema 4.

Teorema 5. *Sea V un espacio de dimensión n sobre el cuerpo de los números reales y sea f una forma bilineal simétrica sobre V que tiene rango r. Entonces existe una base ordenada $\{\beta_1, \beta_2, \ldots, \beta_n\}$ de V en la que la matriz de f es diagonal y tal que*

$$f(\beta_j, \beta_j) = \pm 1, \qquad j = 1, \ldots, r.$$

Además, el número de vectores de base β_j para los que $f(\beta_j, \beta_j) = 1$ es independiente de la elección de base.

Demostración. Existe una base $\{\alpha_1, \ldots, \alpha_n\}$ de V tal que

$$\begin{aligned} f(\alpha_i, \alpha_j) &= 0, & i &\neq j \\ f(\alpha_j, \alpha_j) &\neq 0, & 1 &\leq j \leq r \\ f(\alpha_j, \alpha_j) &= 0, & j &> r. \end{aligned}$$

Sea

$$\begin{aligned} \beta_j &= |f(\alpha_j, \alpha_j)|^{-1/2}\alpha_j, & 1 &\leq j \leq r \\ \beta_j &= \alpha_j, & j &> r. \end{aligned}$$

Entonces $\{\beta_1, \ldots, \beta_n\}$ es una base con las propiedades establecidas.

Sea p el número de vectores de base β_j para los que $f(\beta_j, \beta_j) = 1$; se debe demostrar que el número p es independiente de la base particular que tenemos, que satisface las condiciones establecidas. Sea V^+ el subespacio de V generado por los vectores de base β_j para los que $f(\beta_j, \beta_j) = 1$ y sea V^- el subespacio generado por los vectores de base β_j para los que $f(\beta_j, \beta_j) = -1$. Ahora $p = \dim V^+$, de modo que es la unicidad de la dimensión de V^+ la que se debe demostrar. Es fácil ver que si α es un vector no nulo de V^+, entonces $f(\alpha, \alpha) > 0$; en otras palabras, f es positivamente definida sobre el espacio V^+. En forma análoga, si α es un vector no nulo de V^-, entonces $f(\alpha, \alpha) < 0$; es decir, f es negativamente definida en el subespacio V^-. Sea ahora $V^\perp$ el subespacio generado por los vectores de base β_j para los que $f(\beta_j, \beta_j) = 0$. Si α está en $V^\perp$, entonces $f(\alpha, \beta) = 0$ para todo β en V.

Como $\{\beta_1, \ldots, \beta_n\}$ es una base de V, se tiene que

$$V = V^+ \oplus V^- \oplus V^\perp.$$

Además, se afirma que si W es cualquier subespacio de V en el que f es positivamente definida, entonces los subespacios W, V^- y $V^\perp$ son independientes. En efecto, supóngase que α está en W, β en V^-, γ en V y que $\alpha + \beta + \gamma = 0$. Entonces

$$\begin{aligned} 0 &= f(\alpha, \alpha + \beta + \gamma) = f(\alpha, \alpha) + f(\alpha, \beta) + f(\alpha, \gamma) \\ 0 &= f(\beta, \alpha + \beta + \gamma) = f(\beta, \alpha) + f(\beta, \beta) + f(\beta, \gamma). \end{aligned}$$

Como γ está en $V^\perp$, $f(\alpha, \gamma) = f(\beta, \gamma) = 0$; y como f es simétrica, se obtiene

$$\begin{aligned} 0 &= f(\alpha, \alpha) + f(\alpha, \beta) \\ 0 &= f(\beta, \beta) + f(\alpha, \beta) \end{aligned}$$

Luego $f(\alpha, \alpha) = f(\beta, \beta)$. Como $f(\alpha, \alpha) \geq 0$ y $f(\beta, \beta) < 0$, se sigue que

$$f(\alpha, \alpha) = f(\beta, \beta) = 0.$$

Pero f es positivamente definida sobre W y negativamente definida sobre V^- Se concluye que $\alpha = \beta = 0$ y luego también que $\gamma = 0$. Como

$$V = V^+ \oplus V^- \oplus V^\perp$$

y W, V^-, $V^\perp$ son independientes, vemos que $\dim W \leq \dim V^+$. Esto es, si W es cualquier subespacio de V en el que f es positivamente definida, la di-

mensión de W no puede exceder la dimensión de V^+. Si $\mathcal{B}_1$ es otra base ordenada de V que satisface las condiciones del teorema, se tendrán los correspondientes subespacios V_1^+, V_1^- y $V_1^\perp$; y el razonamiento anterior muestra que $\dim V_1^+ \leq \dim V^+$. Invirtiéndolo, se obtiene que $\dim V^+ \leq \dim V_1^+$ y, por tanto,

$$\dim V^+ = \dim V_1^+. \quad \blacksquare$$

Varios comentarios habría que hacer respecto a la base $\{\beta_1, \ldots, \beta_n\}$ del Teorema 5 y de los subespacios asociados V^+, V^- y $V^\perp$. Primero obsérvese que $V^\perp$ es exactamente el subespacio de los vectores que son «ortogonales» a todos los de V. Se observó anteriormente que $V^\perp$ está contenido en este subespacio; pero

$$\dim V^\perp = \dim V - (\dim V^+ + \dim V^-) = \dim V - \text{rango } f$$

de modo que todo vector α tal que $f(\alpha, \beta) = 0$ para todo β debe estar en $V^\perp$. Así que el subespacio $V^\perp$ es único. Los subespacios V^+ y V^- no son únicos; sin embargo, sus dimensiones son únicas. La demostración del Teorema 5 muestra que V^+ es la mayor dimensión posible de cualquier subespacio en el que f es positivamente definida. En forma análoga, $\dim V^-$ es la mayor dimensión de cualquier subespacio en el que f es negativamente definida. Por cierto

$$\dim V^+ + \dim V^- = \text{rango } f.$$

El número

$$\dim V^+ - \dim V^-$$

es a menudo llamado **signatura** de f. Se le introduce porque las dimensiones de V^+ y V^- están determinadas fácilmente por el rango de f y la signatura de f.

Debería hacerse tal vez un comentario sobre la relación de las formas bilineales simétricas sobre espacios vectoriales reales con los productos internos. Supóngase que V es un espacio vectorial real de dimensión finita y que V_1, V_2, V_3 sean subespacios de V tales que

$$V = V_1 \oplus V_2 \oplus V_3.$$

Supóngase que f_1 es un producto interno sobre V_1 y f_2 es un producto interno sobre V_2. Se puede entonces definir una forma bilineal simétrica f sobre V del siguiente modo: si α, β son vectores de V, entonces se puede escribir

$$\alpha = \alpha_1 + \alpha_2 + \alpha_3 \quad \text{y} \quad \beta = \beta_1 + \beta_2 + \beta_3$$

con α_j, y β_j en V_j. Sea

$$f(\alpha, \beta) = f_1(\alpha_1, \beta_1) - f_2(\alpha_2, \beta_2).$$

El subespacio $V^\perp$ para f será V_3; V_1 es un adecuado V^+ para f, y V_2 es un adecuado V^-. Una parte de la afirmación del Teorema 5 es que toda forma bilineal simétrica sobre V surge de este modo. El resto del contenido del teorema es que un producto interno está representado en alguna base ordenada por la matriz unidad.

Ejercicios

1. Las siguientes expresiones definen formas cuadráticas q sobre R^2. Hallar la forma bilineal simétrica f correspondiente a cada q.

(a) ax_1^2.
(b) bx_1x_2.
(c) cx_2^2.
(d) $2x_1^2 - \frac{1}{3}x_1x_2$.
(e) $x_1^2 + 9x_2^2$.
(f) $3x_1x_2 - x_2^2$.
(g) $4x_1^2 + 6x_1x_2 - 3x_2^2$.

2. Hallar la matriz, en la base ordenada canónica, y el rango de cada una de las formas bilineales determinadas en el Ejercicio 1. Indicar cuáles formas son no degeneradas.

3. Sea $q(x_1, x_2) = ax_1^2 + bx_1x_2 + cx_2^2$ la forma cuadrática asociada con una forma bilineal simétrica f sobre R^2. Demostrar que f es no degenerada si, y solo si, $b^2 - 4ac \neq 0$.

4. Sea V un espacio vectorial de dimensión finita sobre un subcuerpo F de los números complejos y sea S el conjunto de todas las formas bilineales simétricas sobre V.
(a) Demostrar que S es un subespacio de $L(V, V, F)$.
(b) Hallar dim S.

Sea Q el conjunto de todas las formas cuadráticas sobre V.

(c) Demostrar que Q es un subespacio del espacio de todas las funciones de V en F.
(d) Describir explícitamente un isomorfismo T de Q sobre S, sin referencia a una base.
(e) Sea U un operador lineal sobre V y q un elemento de Q. Mostrar que la igualdad $(U^\dagger q)(\alpha) = q(U\alpha)$ define una forma cuadrática $U^\dagger q$ sobre V.
(f) Si U es un operador lineal sobre V, demostrar que la función $U^\dagger$ definida en la parte (e) es un operador lineal sobre Q. Demostrar que $U^\dagger$ es inversible si, y solo si, U es inversible.

5. Sea q la forma cuadrática sobre R^2 dada por

$$q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2, \qquad a \neq 0.$$

Hallar un operador lineal inversible U sobre R^2 tal que

$$(U^\dagger q)(x_1, x_2) = ax_1^2 + \left(c - \frac{b^2}{a}\right)x_2^2.$$

(*Sugerencia*: Para hallar U^{-1} (y luego U) completar el cuadrado. Para la definición de $U^\dagger$ véase parte (e) del Ejercicio 4.)

6. Sea q la forma cuadrática sobre R^2 dada por

$$q(x_1, x_2) = 2bx_1x_2.$$

Hallar un operador lineal inversible U sobre R^2 tal que

$$(U^\dagger q)(x_1, x_2) = 2bx_1^2 - 2bx_2^2.$$

7. Sea q la forma cuadrática sobre R^3 dada por

$$q(x_1, x_2, x_3) = x_1x_2 + 2x_1x_3 + x_3^2.$$

Hallar un operador lineal inversible U sobre R^3 tal que

$$(U^\dagger q)(x_1, x_2, x_3) = x_1^2 - x_2^2 + x_3^2.$$

(*Sugerencia*: Expresar U como producto de operadores análogos a los usados en el Ejercicio 5 y 6.)

8. Sea A una matriz simétrica $n \times n$ sobre R, y sea q la forma cuadrática sobre R^n dada por

$$q(x_1, \ldots, x_n) = \sum_{i,j} A_{ij}x_ix_j.$$

Generalizar el método usado en el Ejercicio 7 para demostrar que existe un operador lineal inversible U sobre R^n tal que

$$(U^\dagger q)(x_1, \ldots, x_n) = \sum_{i=1}^{n} c_ix_i^2$$

donde c_i es 1, -1, o 0, $i = 1, \ldots, n$.

9. Sea f una forma bilineal simétrica sobre R^n. Usando el resultado del Ejercicio 8, demostrar que existe una base ordenada $\mathcal{B}$ tal que $[f]_{\mathcal{B}}$ es diagonal.

10. Sea V el espacio vectorial real de todas las matrices (complejas) hermíticas 2×2, esto es, matrices complejas 2×2, A, que cumplen $A_{ij} = \bar{A}_{ji}$.

(a) Demostrar que la igualdad $q(A) = \det A$ define una forma cuadrática q sobre V.

(b) Sea W el subespacio de V de las matrices de traza 0. Demostrar que la forma bilineal f determinada por q es negativamente definida en el espacio W

11. Sean V un espacio vectorial de dimensión finita y f una forma bilineal simétrica no degenerada sobre V. Demostrar que, para todo operador lineal T sobre V, existe un único operador T' sobre V tal que $f(T\alpha, \beta) = f(\alpha, T'\beta)$ para todo α, β de V. Demostrar también que

$$(T_1T_2)' = T_2'T_1'$$
$$(c_1T_1 + c_2T_2)' = c_1T_1' + c_2T_2'$$
$$(T')' = T.$$

¿Cuánto de lo anterior sigue siendo válido sin la suposición de que T es no degenerada?

12. Sean F un cuerpo y V el espacio de las matrices $n \times 1$ sobre F. Supóngase que A es una matriz $n \times n$ dada sobre F y que f es la forma bilineal sobre V definida por $f(X, Y) = X^tAY$. Supóngase que f es simétrica y no degenerada. Sea B una matriz $n \times n$ sobre F y T el operador lineal sobre V que aplica X en BX. Hallar el operador T' del Ejercicio 11.

13. Sean V un espacio vectorial de dimensión finita y f una forma bilineal simétrica no degenerada sobre V. Asociado a f hay un isomorfismo «natural» de V sobre el espacio dual V^*; este isomorfismo es la transformación L_f de la Sección 10.1. Usando L_f, demostrar que para cada base $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ de V existe una única base $\mathcal{B}' = \{\alpha_1', \ldots, \alpha_n'\}$ de V tal que $f(\alpha_i, \alpha_j') = \delta_{ij}$. Mostrar entonces que para todo vector α de V se tiene que

$$\alpha = \sum_i f(\alpha, \alpha_i')\alpha_i = \sum_i f(\alpha_i, \alpha)\alpha_i'.$$

14. Sean V, f, $\mathcal{B}$ y $\mathcal{B}'$ como en el Ejercicio 13. Supóngase que T es un operador lineal sobre V y que T' es el operador que f asocia a T como en el Ejercicio 11. Demostrar que

(a) $[T']_{\mathcal{B}'} = [T]_{\mathcal{B}}^t$.

(b) $\operatorname{tr}(T) = \operatorname{tr}(T') = \sum_i f(T\alpha_i, \alpha_i')$.

15. Sean V, f, $\mathcal{B}$ y $\mathcal{B}'$ como en el Ejercicio 13. Supóngase que $[f]_{\mathcal{B}} = A$. Demostrar que

$$\alpha_i' = \sum_j (A^{-1})_{ij}\alpha_j = \sum_j (A^{-1})_{ji}\alpha_j.$$

16. Sean F un cuerpo y V el espacio de las matrices $n \times 1$ sobre F. Supóngase que A es una matriz $n \times n$ inversible y simétrica sobre F y que f es la forma bilineal sobre V definida por $f(X, Y) = X^tAY$. Sean P una matriz inversible $n \times n$ sobre F y $\mathcal{B}$ la base de V que consiste en las columnas de P. Hacer ver que la base $\mathcal{B}'$ del Ejercicio 13 consta de las columnas de la matriz $A^{-1}(P^t)^{-1}$.

17. Sean V un espacio vectorial de dimensión finita sobre un cuerpo F y f la forma bilineal simétrica sobre V. Para cada subespacio W de V, sea $W^\perp$ el conjunto de todos los vectores α de V de modo que $f(\alpha, \beta) = 0$ para todo β de W. Demostrar que

(a) $W^\perp$ es un subespacio.
(b) $V = \{0\}^\perp$.
(c) $V^\perp = \{0\}$ si, y solo si, f es no degenerada.
(d) rango $f = \dim V - \dim V^\perp$
(e) Si $\dim V = n$ y $\dim W = m$, entonces $\dim W^\perp \geq n - m$. (Sugerencia: Sea $\{\beta_1, \ldots, \beta_n\}$ una base de W y considérese la aplicación

$$\alpha \to (f(\alpha, \beta_1), \ldots, f(\alpha, \beta_m))$$

de V en F^m).
(f) La restricción de f a W es no degenerada si, y solo si,

$$W \cap W^\perp = \{0\}.$$

(g) $V = W \oplus W^\perp$ si, y solo si, la restriccion de f a W es no degenerada.

18. Sean V un espacio vectorial de dimensión finita sobre C y f una forma bilineal simétrica no degenerada sobre V. Demostrar que existe una base $\mathcal{B}$ de V tal que $\mathcal{B}' = \mathcal{B}$. (Véase Ejercicio 13 para la definición de $\mathcal{B}'$.)

10.3. Formas bilineales antisimétricas

En toda esta sección, V será un espacio vectorial sobre un subcuerpo F del cuerpo de los números complejos. Una forma bilineal f en V se llama **antisimétrica** si $f(\alpha, \beta) = -f(\beta, \alpha)$ para todos los vectores α, β de V. Se demostrará un teorema que se refiere a la simplificación de la matriz de una forma bilineal antisimétrica sobre un espacio de dimensión finita V. Primero, se harán algunas observaciones generales.

Supóngase que f es *cualquier* forma sobre V. Si se hace

$$g(\alpha, \beta) = \tfrac{1}{2}[f(\alpha, \beta) + f(\beta, \alpha)]$$
$$h(\alpha, \beta) = \tfrac{1}{2}[f(\alpha, \beta) - f(\beta, \alpha)]$$

entonces es fácil verificar que g es una forma bilineal simétrica sobre V y que h es una forma bilineal antisimétrica sobre V. También que $f = g + h$. Además, esta expresión de V, como suma de una forma simétrica y una antisimétrica, es única. Así el espacio $L(V, V, F)$ es suma directa del subespacio de las formas simétricas y del subespacio de las formas antisimétricas.

Si V es de dimensión finita, la forma bilineal f es antisimétrica si, y solo si, su matriz A en alguna (o en toda) base ordenada es antisimétrica, $A^t = -A$. Esto se demuestra del mismo modo que se demostró el correspondiente hecho respecto a las formas bilineales simétricas. Cuando f es antisimétrica, la matriz de f en cualquier base ordenada tendrá todos sus elementos en la diagonal iguales a 0. Esto corresponde a la observación de que $f(\alpha, \alpha) = 0$ para todo α de V, ya que $f(\alpha, \alpha) = -f(\alpha, \alpha)$.

Supóngase que f es una forma bilineal antisimétrica no nula sobre V. Como $f \neq 0$, existen vectores α, β de V tales que $f(\alpha, \beta) \neq 0$. Multiplicando α por un escalar apropiado, se puede suponer que $f(\alpha, \beta) = 1$. Sea γ cualquier vector del subespacio generado por α y β, por ejemplo, $\gamma = c\alpha + d\beta$. Entonces

$$\begin{aligned} f(\gamma, \alpha) &= f(c\alpha + d\beta, \alpha) = df(\beta, \alpha) = -d \\ f(\gamma, \beta) &= f(c\alpha + d\beta, \beta) = cf(\alpha, \beta) = c \end{aligned}$$

y así

(10-7) $$\gamma = f(\gamma, \beta)\alpha - f(\gamma, \alpha)\beta.$$

En particular, obsérvese que α y β son necesariamente independientes; en efecto, si $\gamma = 0$, entonces $f(\gamma, \alpha) = f(\gamma, \beta) = 0$.

Sea W el espacio bidimensional generado por α y β. Sea $W^\perp$ el conjunto de todos los vectores δ de V tales que $f(\delta, \alpha) = f(\delta, \beta) = 0$; esto es, el conjunto de todos los δ tales que $f(\delta, \gamma) = 0$ para todo γ en el subespacio W. Se afirma que $V = W \oplus W^\perp$. En efecto, sea ϵ cualquier vector de V y

$$\begin{aligned} \gamma &= f(\epsilon, \beta)\alpha - f(\epsilon, \alpha)\beta \\ \delta &= \epsilon - \gamma. \end{aligned}$$

Entonces γ está en W, y δ está en $W^\perp$, para

$$\begin{aligned} f(\delta, \alpha) &= f(\epsilon - f(\epsilon, \beta)\alpha + f(\epsilon, \alpha)\beta, \alpha) \\ &= f(\epsilon, \alpha) + f(\epsilon, \alpha)f(\beta, \alpha) \\ &= 0 \end{aligned}$$

y en forma análoga, $f(\delta, \beta) = 0$. Así, pues, todo ϵ de V es de la forma $\epsilon = \gamma + \delta$ con γ en W y δ en $W^\perp$. De (9-7) es claro que $W \cap W^\perp = \{0\}$, y así $V = W \oplus W^\perp$.

Ahora la restricción de f a $W^\perp$ es una forma bilineal antisimétrica sobre $W^\perp$. Esta restricción puede ser la forma cero. Si no es éste el caso, existen vectores α' y β' en $W^\perp$ tales que $f(\alpha', \beta') = 1$. Si W' es el subespacio bidimensional generado por α' y β', entonces se tendrá que

$$V = W \oplus W' \oplus W_0$$

donde W_0 es el conjunto de todos los vectores δ de $W^\perp$ tales que $f(\alpha', \delta) = f(\beta', \delta) = 0$. Si la restricción de f a W_0 no es la forma cero, se pueden elegir vectores α'', β'' en W_0 tales que $f(\alpha'', \beta'') = 1$, y se sigue así.

En el caso de dimensión finita, se obtiene inmediatamente una sucesión finita de parejas de vectores

$$(\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_k, \beta_k)$$

con las siguientes propiedades:

(a) $f(\alpha_j, \beta_j) = 1$, $j = 1, \ldots, k$.
(b) $f(\alpha_i, \alpha_j) = f(\beta_i, \beta_j) = f(\alpha_i, \beta_j) = 0$, $i \neq j$.
(c) Si W_j es el subespacio bidimensional generado por α_j y β_j, entonces

$$V = W_1 \oplus \cdots \oplus W_k \oplus W_0$$

donde todo vector de W_0 es «ortogonal» a todos los α_j y β_j, y la restricción de f a W_0 es la forma cero.

Teorema 6. *Sea V un espacio vectorial de dimensión n sobre un subcuerpo de los números complejos y sea f una forma bilineal antisimétrica sobre V. Entonces el rango r de f es par, y si $r = 2k$, existe una base ordenada de V en que la matriz de f es suma directa de la matriz cero $(n - r) \times (n - r)$ y k matrices 2×2 iguales a*

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Demostración. Sean $\alpha_1, \beta_1, \ldots, \alpha_k, \beta_k$ los vectores que satisfacen las condiciones anteriores (a), (b) y (c). Sea $\{\gamma_1, \ldots, \gamma_s\}$ cualquier base ordenada del subespacio W_0. Entonces

$$\mathfrak{B} = \{\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_k, \beta_k, \gamma_1, \ldots, \gamma_s\}$$

es una base ordenada de V. Por (a), (b) y (c) es inmediato que la matriz de f en la base ordenada $\mathfrak{B}$ es suma directa de la matriz cero $(n - 2k) \times (n - 2k)$ y k matrices 2×2 iguales a

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{10-8}$$

Además, es claro que el rango de esta matriz y, por tanto, el de f es $2k$. ∎

Una consecuencia de lo anterior es que si f es una forma bilineal antisimétrica no degenerada sobre V, entonces la dimensión de V debe ser par. Si $\dim V = 2k$, existirá una base ordenada $\{\alpha_1, \beta_1, \ldots, \alpha_k, \beta_k\}$ de V tal que

$$f(\alpha_i, \beta_j) = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

$$f(\alpha_i, \alpha_j) = f(\beta_i, \beta_j) = 0.$$

La matriz f en esta base ordenada es suma directa de k matrices antisimétricas 2×2 como la (10-8). Se obtiene otra forma canónica para la matriz de una forma antisimétrica no degenerada si, en vez de la base ordenada dada anteriormente, se considera la base ordenada

$$\{\alpha_1, \ldots, \alpha_k, \beta_k, \ldots, \beta_1\}.$$

El lector hallará fácil verificar que la matriz de f en la última base ordenada tiene la forma bloque

que «preservan» f. El conjunto de todas las matrices $[T]_{\mathcal{B}}$, donde T es un operador lineal que preserva f, será grupo para la multiplicación matricial. Hay otra descripción de este grupo de matrices, y es la siguiente. Sea $A = [f]_{\mathcal{B}}$, de modo que si α y β son vectores de V con las respectivas matrices de coordenadas X e Y respecto de $\mathcal{B}$, se tendrá que

$$f(\alpha, \beta) = X^tAY.$$

Sea T cualquier operador lineal sobre V y $M = [T]_{\mathcal{B}}$. Entonces

$$\begin{aligned} f(T\alpha, T\beta) &= (MX)^tA(MY) \\ &= X^t(M^tAM)Y. \end{aligned}$$

Por tanto, T preserva f si, y solo si, $M^tAM = A$. Entonces, en lenguaje matricial, el Teorema 7 dice lo siguiente: si A es una matriz $n \times n$ inversible, el conjunto de todas las matrices $n \times n$, M, tales que $M^tAM = A$ es grupo para la multiplicación matricial. Si $A = [f]_{\mathcal{B}}$, entonces M está en este grupo de matrices si, y solo si, $M = [T]_{\mathcal{B}}$, donde T es un operador lineal que preserva f.

Antes de ver algunos ejemplos queremos hacer una observación más. Supóngase que f es una forma bilineal simétrica. Un operador lineal T preserva f si, y solo si, T preserva la forma cuadrática

$$q(\alpha) = f(\alpha, \alpha)$$

asociada a f. Si T preserva f, ciertamente tenemos que

$$q(T\alpha) = f(T\alpha, T\alpha) = f(\alpha, \alpha) = q(\alpha)$$

para todo α de V. Recíprocamente, como f es simétrica, la identidad de polarización

$$f(\alpha, \beta) = \tfrac{1}{4}q(\alpha + \beta) - \tfrac{1}{4}q(\alpha - \beta)$$

muestra que T preserva f con tal que $q(T\gamma) = q(\gamma)$ para todo γ de V. (Se está suponiendo aquí que el cuerpo escalar es un subcuerpo de los números complejos.)

Ejemplo 6. Sea V el espacio R^n o C^n. Sea f la forma bilineal

$$f(\alpha, \beta) = \sum_{j=1}^{n} x_jy_j$$

donde $\alpha = (x_1, \ldots, x_n)$ y $\beta = (y_1, \ldots, y_n)$. El grupo que preserva f es llamado **grupo ortogonal** (real o complejo) n-dimensional. El nombre de «grupo ortogonal» más corrientemente se aplica al grupo asociado de matrices en la base ordenada canónica. Como la matriz de f en la base canónica es I, este grupo consiste en las matrices M que cumplen $M^tM = I$. Una matriz así se llama **matriz ortogonal** $n \times n$ (real o compleja). Los dos grupos ortogonales $n \times n$ se designan corrientemente por $O(n, R)$ y $O(n, C)$. Naturalmente, el grupo ortogonal es también el grupo que preserva la forma cuadrática

$$q(x_1, \ldots, x_n) = x_1^2 + \cdots + x_n^2.$$

Ejemplo 7. Sea f la forma bilineal simétrica sobre R^n con forma cuadrática

$$q(x_1, \ldots, x_n) = \sum_{j=1}^{p} x_j^2 - \sum_{j=p+1}^{n} x_j^2.$$

Entonces f es no degenerada y tiene signatura $2p - n$. El grupo de matrices que preservan una forma de este tipo se llama **grupo seudoortogonal.** Cuando $p = n$ se tiene el grupo ortogonal $O(n, R)$ como un tipo particular de grupo seudoortogonal. Para cada uno de los $n + 1$ valores $p = 0, 1, 2, \ldots, n$ se obtienen formas bilineales f diferentes; sin embargo, para $p = k$ y $p = n - k$ las formas son opuestas una de otra y tienen, pues, el mismo grupo asociado. Así, cuando n es impar, se tienen $(n + 1)/2$ grupos seudoortogonales de matrices $n \times n$, y cuando n es par tenemos $(n + 2)/2$ de tales grupos.

Teorema 8. *Sea V un espacio vectorial de dimensión n sobre el cuerpo de los números complejos y sea f una forma bilineal simétrica no degenerada sobre V. Entonces el grupo que preserva f es isomorfo al grupo ortogonal complejo $O(n, C)$.*

Demostración. Por supuesto, un isomorfismo entre dos grupos significa correspondencia biunívoca entre sus elementos que «preservan» la operación de grupo. Sea G el grupo de los operadores lineales sobre V que preservan la forma bilineal f. Como f es simétrica y no degenerada, el Teorema 4 dice que existe una base ordenada $\mathcal{B}$ de V en la que f está representada por la matriz identidad $n \times n$. Por tanto, un operador lineal T preserva f si, y solo si, su matriz en la base ordenada $\mathcal{B}$ es una matriz ortogonal compleja. Luego

$$T \rightarrow [T]_{\mathcal{B}}$$

es un isomorfismo de G sobre $O(n, C)$.

Teorema 9. *Sea V un espacio vectorial de dimensión n sobre el cuerpo de los números reales y sea f una forma bilineal simétrica no degenerada sobre V. Entonces el grupo que preserva f es isomorfo a un grupo seudoortogonal $n \times n$.*

Demostración. Repetir la demostración del Teorema 8, usando ahora el Teorema 5 en vez del 4. ▌

Ejemplo 8. Sea f la forma bilineal simétrica sobre R^4 con la forma cuadrática

$$q(x, y, z, t) = t^2 - x^2 - y^2 - z^2.$$

Un operador lineal T sobre R^4 que preserva esta forma bilineal (o cuadrática) particular, se llama **transformación de Lorentz,** y el grupo que preserva f se llama **grupo de Lorentz.** He aquí un método para describir algunas transformaciones de Lorentz.

Sea H el espacio vectorial real de las matrices complejas 2×2, A, hermíticas, $A = A^*$. Es de fácil verificación que

$$\Phi(x, y, z, t) = \begin{bmatrix} t + x & y + iz \\ y - iz & t - x \end{bmatrix}$$

define un isomorfismo Φ de R^4 sobre el espacio H. Por este isomorfismo la forma cuadrática q se aplica sobre la función determinante, esto es

$$q(x, y, z, t) = \det \begin{bmatrix} t + x & y + iz \\ y - iz & t - x \end{bmatrix}$$

o

$$q(\alpha) = \det \Phi(\alpha).$$

Lo cual sugiere que se pueden estudiar las transformaciones de Lorentz sobre R^4 estudiando los operadores lineales sobre H que preservan determinantes.

Sea M una matriz compleja 2×2, y para una matriz hermítica A, defínase

$$U_M(A) = MAM^*.$$

Ahora bien, MAM^* es también hermítica. Partiendo de esto es fácil ver que U_M es un operador lineal (real) sobre H. Veamos cuándo es cierto que U_M «preserva» determinantes; es decir, $\det U_M(A) = \det A$ para toda A de H. Como el determinante de M^* es el complejo conjugado del determinante de M, se ve que

$$\det [U_M(A)] = |\det M|^2 \det A.$$

Así que U_M preserva determinantes precisamente cuando $\det M$ tiene valor absoluto 1.

Así que tomando ahora una matriz compleja 2×2, M, para la que $|\det M| = 1$, entonces U_M es un operador linea sobre H que preserva determinantes. Se define

$$T_M = \Phi^{-1} U_M \Phi.$$

Como Φ es un isomorfismo, T_M es un operador lineal sobre R^4. También, T_M es una transformación de Lorentz; en efecto,

$$\begin{aligned} q(T_M\alpha) &= q(\Phi^{-1}U_M\Phi\alpha) \\ &= \det (\Phi\Phi^{-1}U_M\Phi\alpha) \\ &= \det (U_M\Phi\alpha) \\ &= \det (\Phi\alpha) \\ &= q(\alpha) \end{aligned}$$

y de este modo T_M preserva la forma cuadrática q. Utilizando matrices 2×2, M, específicas, se puede emplear el método anterior para calcular transformaciones de Lorentz específicas. Dos comentarios al respecto, que no son difíciles de verificar.

(1) Si M_1 y M_2 son matrices 2×2, inversibles con elementos complejos, entonces $U_{M_1} = U_{M_2}$ si, y solo si, M_2 es un múltiplo escalar de M_1. Así, todo lo de las transformaciones de Lorentz vistas anteriormente se puede obtener de las matrices unimodulares M, esto es, de matrices M que satisfacen

det $M = 1$. Si M_1 y M_2 son matrices unimodulares tales que $M_1 \neq M_2$ y $M_1 \neq -M_2$, entonces $T_{M_1} \neq T_{M_2}$.

(2) No toda transformación de Lorentz se puede obtener por el método anterior.

Ejercicios

1. Sea M un elemento del grupo ortogonal complejo $O(n, C)$. Mostrar que M^t, $\bar{M}$ y $M^* = \bar{M}^t$ también pertenecen a $O(n, C)$.

2. Supóngase que M pertenece a $O(n, C)$ y que M' es semejante a M. ¿Pertenece M' también a $O(n, C)$?

3. Sea

$$y_j = \sum_{k=1}^{n} M_{jk}x_k$$

donde M es un elemento de $O(n, C)$. Demostrar que

$$\sum_j y_j^2 = \sum_j x_j^2.$$

4. Sea M una matriz $n \times n$ sobre C^n con columnas $M_1, M_2, \ldots, M_n$. Demostrar que M pertenece a $O(n, C)$ si, y solo si,

$$M_j^t M_k = \delta_{jk}.$$

5. Sea X una matriz $n \times 1$ sobre C. ¿En qué condiciones $O(n, C)$ contiene una matriz M cuya primera columna es X?

6. Hallar una matriz en $O(3, C)$ cuya primera fila es $(2i, 2i, 3)$.

7. Sean V el espacio de todas las matrices $n \times 1$ sobre C y f la forma bilineal sobre V dada por $f(X, Y) = X^tY$. Sea M perteneciente a $O(n, C)$. ¿Cuál es la matriz de f en la base de V formada por las columnas $M_1, M_2, \ldots, M_n$ de M?

8. Sean X una matriz $n \times 1$ sobre C tal que $X^tX = 1$ e Y_j la j-ésima columna de la matriz unidad. Demostrar que existe una matriz M en $O(n, C)$ tal que $MX = I_j$. Si X tiene elementos reales, mostrar que hay una M en $O(n, C)$ con la propiedad de que $MX = I_j$.

9. Sean V el espacio de todas las matrices $n \times 1$ sobre C, A una matriz $n \times n$ sobre C y f la forma bilineal sobre V dada por $f(X, Y) = X^tAY$. Demostrar que f es invariante por $O(n, C)$; es decir, $f(MX, MY) = f(X, Y)$ para todo X, Y de V y M en $O(n, C)$ si, y solo si, A conmuta con cada elemento de $O(n, C)$.

10. Sean S cualquier conjunto de matrices $n \times n$ sobre C y S' el conjunto de todas las matrices $n \times n$ sobre C que conmutan con cada elemento de S. Demostrar que S' es un álgebra sobre C.

11. Sean F un subcuerpo de C, V un espacio vectorial de dimensión finita sobre F y f una forma bilineal no singular sobre V. Si T es un operador lineal sobre V que preserva f, demostrar que det $T = \pm 1$.

12. Sean F un subcuerpo de C, V el espacio de las matrices $n \times 1$ sobre F, A una matriz $n \times n$ inversible sobre F y f una forma bilineal sobre V dada por $f(X, Y) = X^tAY$. Si M es una matriz $n \times n$ sobre F, demostrar que M preserva f si, y solo si, $A^{-1}M^tA = M^{-1}$.

13. Sea g una forma bilineal no singular sobre un espacio vectorial de dimensión finita V. Supóngase que T es un operador lineal inversible sobre V y que f es la forma bilineal sobre V dada por $f(\alpha, \beta) = g(\alpha, T\beta)$. Si U es un operador lineal sobre V, hallar condiciones necesarias y suficientes de U para preservar f.

14. Sea T un operador lineal sobre C^2 que preserva la forma cuadrática $x_1^2 - x_2^2$. Demostrar que

(a) $\det (T) = \pm 1$.

(b) Si M es la matriz de T en la base canónica, entonces $M_{22} = \pm M_{11}$, $M_{21} = \pm M_{12}$, $M_{11}^2 - M_{11}^2 = 1$.

(c) Si $\det M = 1$, entonces existe un número complejo no nulo c tal que

$$M = \frac{1}{2}\begin{bmatrix} c + \frac{1}{c} & c - \frac{1}{c} \\ c - \frac{1}{c} & c + \frac{1}{c} \end{bmatrix}.$$

(d) Si $\det M = -1$, entonces existe un número complejo c tal que

$$M = \frac{1}{2}\begin{bmatrix} c + \frac{1}{c} & c - \frac{1}{c} \\ -c + \frac{1}{c} & -c - \frac{1}{c} \end{bmatrix}.$$

15. Sea f la forma bilineal sobre C^2 definida por

$$f((x_1, x_2), (y_1, y_2)) = x_1 y_2 - x_2 y_1.$$

Demostrar que

(a) Si T es un operador lineal sobre C^2, entonces $f(T\alpha, T\beta) = (\det T)f(\alpha, \beta)$ para todo α, β en C^2.

(b) T preserva f si y solo si, $\det T = \pm 1$.

(c) ¿Qué dice (b) respecto al grupo de matrices 2×2, M, tales que $M^t A M = A$, donde

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}?$$

16. Sean un entero positivo n, I la matriz unidad $n \times n$ sobre C y J la matriz $2n \times 2n$ dada por

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

Sea M una matriz $2n \times 2n$ sobre C de la forma

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

con A, B, C, D matrices $n \times n$ sobre C. Hallar condiciones necesarias y suficientes de A, B, C, D para que $M^t J M = J$.

17. Hallar todas las formas bilineales sobre el espacio de las matrices $n \times 1$ sobre R que son invariantes por $O(n, R)$.

18. Hallar todas las formas bilineales sobre el espacio de las matrices $n \times 1$, sobre C que son invariantes por $O(n, C)$.

Apéndice

Este apéndice se divide lógicamente en dos partes. La primera, que abarca las primeras tres secciones, contiene ciertos conceptos fundamentales que aparecen a lo largo del libro (en realidad, a lo largo de la matemática). Es más bien una introducción para el libro que un apéndice. La segunda parte es más genuinamente un apéndice al texto.

La Sección 1 contiene un estudio sobre los conjuntos, sus uniones e intersecciones. La Sección 2 analiza el concepto de función y las ideas relacionadas de imagen, dominio, función recíproca y restricción de una función a un subconjunto de su dominio. La Sección 3 trata de las relaciones de equivalencia. La materia en estas tres secciones, especialmente la de las Secciones 1 y 2, está presentada de modo un tanto conciso. Se trata más bien de un convenio sobre terminología que de una exposición detallada. En estricto sentido lógico, este material constituye una parte de los requisitos para la lectura del libro; sin embargo, el lector no debe descorazonarse si no capta completamente el significado de las ideas en su primera lectura. Estas ideas son importantes, pero al lector que no esté del todo familiarizado con ellas le será más fácil absorberlas si revisa la exposición de vez en cuando, mientras lee el texto propiamente.

Las Secciones 4 y 5 se refieren a las relaciones de equivalencia en el contexto del álgebra lineal. La Sección 4 contiene un breve estudio de los espacios cocientes. Puede ser leído en cualquier oportunidad, después de los primeros dos o tres capítulos del libro. La Sección 5 trata brevemente algunas de las relaciones de equivalencia que se presentan en el libro y pretende mostrar cómo algunos de los resultados del libro se pueden interpretar desde el punto de vista de las relaciones de equivalencia. La Sección 6 describe el axioma de elección y sus implicaciones para el álgebra lineal.

A.1. Conjuntos

Usaremos las expresiones «conjunto», «clase», «colección» y «familia» indistintamente, aunque damos preferencia a «conjunto». Si S es un conjunto y x es un objeto del conjunto S, se dirá que x es un **miembro de** S, que x es un **elemento de** S, que x **pertenece a** S o simplemente que x está en S. Si S tiene solo un número finito de elementos, $x_1, \ldots, x_n$, se suele enunciar S disponiendo sus elementos entre paréntesis:

$$S = \{x_1, \ldots, x_n\}.$$

Así el conjunto S de los enteros positivos de 1 a 5 será

$$S = \{1, 2, 3, 4, 5\}.$$

Si S y T son conjuntos, diremos que S es un **subconjunto de** T o que S está **contenido en** T, si todo elemento de S es un elemento de T. Todo conjunto S es subconjunto de sí mismo. Si S es subconjunto de T, pero S y T no son idénticos, diremos que S es **subconjunto propio** de T. En otras palabras, S es suoconjunto propio de T siempre que S esté contenido en T, pero T no esté contenido en S.

Si S y T son conjuntos, la **unión de** S **y** T es el conjunto $S \cup T$ que consta de todos los objetos x que son elementos de S o T. La **intersección de** S **y** T es el conjunto $S \cap T$, que consta de todos los x que son elementos de S y de T. Para dos conjuntos cualesquiera, S y T, la intersección $S \cap T$ es un subconjunto de la unión $S \cup T$. Esto debe contribuir a aclarar el uso de la palabra «o» que prevalecerá en este libro. Cuando se dice que x está en S o en T, no se excluye la posibilidad de que x esté en ambos, S y T.

Para que la intersección de S y T sea siempre un conjunto es necesario que se introduzca el **conjunto vacío,** es decir, el conjunto sin elementos. Entonces $S \cap T$ es el conjunto vacío si, y solo si, S y T no tienen elementos en común.

A menudo se necesitará considerar la unión o intersección de varios conjuntos. Si $S_1, \ldots, S_n$ son conjuntos, su **unión** es el conjunto $\bigcup_{j=1}^{n} S_j$ que consta de todos los x que son elementos de al menos uno de los conjuntos $S_1, \ldots, S_n$. Su **intersección** es el conjunto $\bigcap_{j=1}^{n} S_j$ que consta de todos los x que son elementos de cada uno de los conjuntos $S_1, \ldots, S_n$. En unas pocas ocasiones tendremos que ver con la unión o intersección de una colección infinita de conjuntos. Debería quedar claro cómo tales uniones e intersecciones están definidas. El siguiente ejemplo aclarará estas definiciones y una notación para ellas.

Ejemplo 1. Sea R el conjunto de todos los números reales (el eje real). Si t está en R, se asocia con t un subconjunto S_t de R definido de la siguiente forma: S_t consta de todos los números reales x que no son menores que t.

(a) $S_{t_1} \cup S_{t_2} = S_t$, donde t es el menor entre t_1 y t_2.

(b) $S_{t_1} \cap S_{t_2} = S_t$, donde t es el mayor entre t_1 y t_2.

(c) Sea I el intervalo unidad, esto es, el conjunto de todos los t en R para los que $0 \leq t \leq 1$. Entonces

$$\bigcup_{t \text{ en } I} S_t = S_0$$

$$\bigcap_{t \text{ en } I} S_t = S_1.$$

A.2. Funciones

Una **función** consta de lo siguiente:

(1) un conjunto X, llamado dominio de la función;
(2) un conjunto Y, llamado codominio de la función;
(3) una relación (o correspondencia) f, que asocia a cada elemento x de X un solo elemento $f(x)$ de Y.

Si (X, Y, f) es una función, se dice también que f **es una función de** X **en** Y. Esto no es del todo correcto, ya que no es f quien es la función; f es la ley de la función. Sin embargo, este uso del mismo símbolo para la función y su ley proporciona un medio mucho más manejable para hablar sobre funciones. Así se dirá que f es una función de X en Y, que X es el dominio de f y que Y es el codominio de f—todo esto quiere decir que (X, Y, f) es una función como se definió anteriormente. Existen otras expresiones que comúnmente se usarán en lugar de «función». Algunas de éstas son «transformación», «operador» y «aplicación». Tales expresiones se usarán en contextos donde sean más apropiadas en consideración al papel que debe desempeñar una función particular.

Si f es una función de X en Y, el **conjunto de imágenes** (o **imagen**) de f es el conjunto de todos los $f(x)$, con x de X. En otras palabras, la imagen de f consta de todos los elementos y de Y tales que $y = f(x)$ para algún x de X. Si la imagen de f es todo Y, se dirá que f es una **función de** X **sobre** Y o simplemente que f es **sobreyectiva.** La imagen de f se representa a menudo por $f(X)$.

Ejemplo 2. (a) Sea X el conjunto de los números reales y sea $Y = X$. Sea f la función de X en Y definida por $f(x) = x^2$. La imagen de f es el conjunto de todos los números reales no negativos. Así que f no es sobreyectiva.

(b) Sea X el plano euclidiano, e $Y = X$. Defínase f como sigue: si P es un punto del plano, entonces $f(P)$ es el punto obtenido por rotación de P en 90 (en torno al origen, en la dirección antirreloj). El recorrido de f es todo Y, es decir, el plano entero, y así f es sobreyectiva.

(c) Sea otra vez X el plano euclidiano. Se introduce un sistema de coordenadas en X, como en geometría analítica, por medio de dos rectas perpendiculares para identificar los puntos de X con pares ordenados (x_1, x_2) de números reales. Sea Y el eje x_1, esto es, todos los puntos (x_1, x_2) con $x_2 = 0$. Si P es un punto de X, sea $f(P)$ el punto obtenido al proyectar P sobre el eje x_1, paralelamente al eje x_2. Es decir, $f(x_1, x_2) = (x_1, 0)$. La imagen de f es todo Y, por lo que f es sobreyectiva.

(d) Sea X el conjunto de los números reales y sea Y el conjunto de los números reales positivos. Se define una función f de X en Y por $f(x) = e^x$. Entonces f es una función de X sobre Y.

(e) Sean X el conjunto de los números reales positivos e Y el conjunto de los números reales. Sea f la función logaritmo natural, esto es, la función definida por $f(x) = \log x = \ln x$. Aquí también f es sobreyectiva, es decir, todo número real es el logaritmo natural de algún número positivo.

Supóngase que X, Y y Z sean conjuntos, que f sea una función de X en Y y que g sea una función de Y en Z. Existe una función $g \circ f$ de X en Z asociada con f y g, llamada **composición** de g y f. Está definida por

$$(g \circ f)(x) = g(f(x)).$$

Como un ejemplo simple, sea $X = Y = Z$ el conjunto de los números reales, sean f, g, h las funciones de X en Y definidas por

$$f(x) = x^2, \qquad g(x) = e^x, \qquad h(x) = e^{x^2}$$

y entonces $h = g \circ f$. La composición $g \circ f$ se representa a menudo simplemente por gf; sin embargo, como indica el ejemplo anterior, hay veces en que esto se presta a confusiones.

Un problema de interés es el siguiente. Supóngase que f es una función de X en Y. ¿Cuándo existe una función g de Y en X tal que $g(f(x)) = x$ para todo x de X? Si se representa por I la **función identidad** en X, esto es, la función de X en X definida por $I(x) = x$, se pregunta: ¿cuándo existe una función g de Y en X tal que $g \circ f = I$? En forma somera, se quiere una función g que «envíe cada elemento de Y al sitio de donde proviene». Para que tal función g exista, f debe ser evidentemente inyectiva, esto es, f debe tener la propiedad de que si $x_1 \neq x_2$, entonces $f(x_1) \neq f(x_2)$. Si f es inyectiva, tal función g existe. Se la define como sigue: sea y un elemento de Y. Si y está en la imagen de f, entonces existe un elemento x de X tal que $y = f(x)$; y como f es inyectiva, existe exactamente un x con esa propiedad. Se define $g(y) = x$. Si y no está en la imagen de f se define $g(y)$ como cualquier elemento de X. Obviamente tenemos entonces que $g \circ f = I$.

Sea f una función X en Y. Se dice que f es **inversible** si existe una función g de Y en X tal que

(1) $g \circ f$ es la función identidad en X,
(2) $f \circ g$ es la función identidad en Y.

Acabamos de ver que si existe una función g que satisface (1), entonces f es inyectiva. En forma semejante se puede ver que si existe una g que satisface (2), la imagen de f es todo Y, es decir, f es sobreyectiva. Así, si f es inversible, f es inyectiva y sobreyectiva. Recíprocamente, si f es inyectiva y sobreyectiva, existe una función g de Y en X que satisface (1) y (2). Además, esta g es única. Es la función de Y en X definida por esta ley; si y está en Y, entonces $g(y)$ es el único elemento x en X para el que $f(x) = y$.

Si f es inversible (inyectiva y sobreyectiva), la **recíproca** de f es la función única f^{-1} de Y en X que cumple:

(1′) $f^{-1}(f(x)) = x$ para cada x en X,
(2′) $f(f^{-1}(y)) = y$ para cada y en Y.

Ejemplo 3. Considérense las funciones del Ejemplo 2.

(a) Si $X = Y$, el conjunto de los números reales, y $f(x) = x^2$, entonces f no es inversible. En efecto, f no es inyectiva ni sobreyectiva.

(b) Si $X = Y$, el plano euclidiano, y f es la «rotación de 90°», entonces f es inyectiva y sobreyectiva. La función recíproca f^{-1} es la «rotación de $-90°$» o la «rotación de 270°».

(c) Si X es el plano, Y el eje x_1, y $f(x_1, x_2) = (x_1, 0)$, entonces f no es inversible. En efecto, f es sobreyectiva, pero f no es inyectiva.

(d) Si X es el conjunto de los números reales, Y el conjunto de los números reales positivos y $f(x) = e^x$, entonces f es inversible. La función f^{-1} es la función logaritmo natural de la parte (e): $\log e^x = x$, $e^{\log y} = y$.

(e) La recíproca de esta función logaritmo natural es la función exponencial de la parte (d).

Sea f una función de X en Y y sea f_0 una función de X_0 en Y_0. Se dice que f_0 es una **restricción** de f (o una restricción de f a X_0) si

(1) X_0 es un subconjunto de X,
(2) $f_0(x) = f(x)$ para todo x de X_0.

Por supuesto, que cuando f_0 es una restricción de f, se sigue que Y_0 es un subconjunto de Y. El nombre de «restricción» viene del hecho de que f y f_0 tienen la misma ley, y difieren principalmente porque se ha restringido el dominio de definición de la ley a un subconjunto X_0 de X.

Si hemos dado la función f y cualquier subconjunto X_0 de X, hay una forma obvia para construir una restricción de f a X_0. Definimos una función f_0 de X_0 en Y por $f_0(x) = f(x)$ para todo x de X_0. Podríamos preguntarnos por qué no decimos que esta es *la* restricción de f a X_0. La razón es que cuando se discuten las restricciones de f se desea tener libertad para cambiar el codominio Y, al igual que el dominio X.

Ejemplo 4. (a) Sea X el conjunto de los números reales y f la función de X en X definida por $f(x) = x^2$. Entonces f no es una función inversible, pero lo es si se restringe su dominio al de los números reales no negativos. Sea X_0 el conjunto de los números reales no negativos y sea f_0 la función de X_0 en X_0 definida por $f_0(x) = x^2$. Entonces f_0 es una restricción de f a X_0. Pero f no es inyectiva ni sobreyectiva, mientras que f_0 es inyectiva y sobreyectiva. La última afirmación solo dice que todo número no negativo es raíz cuadrada de un número no negativo precisamente. La función recíproca f_0^{-1} es la función de X_0 en X_0 definida por $f_0^{-1}(x) = \sqrt{x}$.

(b) Sea X el conjunto de los números reales y sea f la función de X en X definida por $f(x) = x^3 + x^2 + 1$. La imagen de f es todo X y f es sobreyectiva.

La función f obviamente no es inyectiva, v.gr., $f(-1) = f(0)$. Pero f es inyectiva en X_0, conjunto de los números reales no negativos, ya que la derivada de f es positiva para $x > 0$. Cuando x recorre todos los números no negativos, $f(x)$ recorre todos los números reales y tales que $y \geq 1$. Si se hace que Y_0 sea el conjunto de todos los $y \geq 1$ y que f_0 sea la función de X_0 en Y_0, definida por $f_0(x) = f(x)$, entonces f_0 es una función inyectiva de X_0 sobre Y_0. Consecuentemente, f_0 tiene una función recíproca f_0^{-1} de Y_0 en X_0. Cualquier fórmula para $f_0^{-1}(y)$ es bastante complicada.

(c) Sea otra vez X el conjunto de los números reales y sea f la función seno, esto es, la función de X en X definida por $f(x) = \text{sen}\, x$. La imagen de f es el conjunto de todos los y tales que $-1 \leq y \leq 1$; luego f no es sobreyectiva. Como $f(x + 2\pi) = f(x)$, vemos que f no es inyectiva. Si se hace que X_0 sea el intervalo $-\pi/2 \leq x \leq \pi/2$, entonces f es inyectiva en X_0. Sea Y_0 el intervalo $-1 \leq y \leq 1$ y sea f_0 la función de X_0 en Y_0 definida por $f_0(x) = \text{sen}\, x$. Entonces f_0 es una restricción de f en el intervalo X_0 y f_0 es inyectiva y sobreyectiva. Esta es otra forma de decir que, en el intervalo de $-\pi/2$ a π 2, la función seno toma todo valor entre -1 y 1 exactamente una vez. La función f_0^{-1} es la función recíproca del seno:

$$f_0^{-1}(y) = \text{sen}^{-1}\, y = \text{arc sen}\, y.$$

(d) Este es un ejemplo general de restricción de una función. Es un ejemplo mucho más característico de los tipos de restricciones que se usarán en este libro, que los anteriores ejemplos (b) y (c). El ejemplo (a) es un caso especial de éste. Sea X un conjunto y f una función de X en sí mismo. Sea X_0 un subconjunto de X. Se dice que X_0 es **invariante por** f si para todo x de X_0 el elemento $f(x)$ está en X_0. Si X_0 es invariante por f, entonces f induce una función f_0 de X_0 en sí mismo, por restricción del dominio de su definición a X_0. La importancia de la invariancia es que por restricción de f a X_0 se puede obtener una función de X_0 en sí mismo, más bien que simplemente una función de X_0 en X.

A.3. *Relaciones de equivalencia*

Una relación de equivalencia es un tipo específico de relación entre pares ordenados de elementos en un conjunto. Para definir una relación de equivalencia debemos ver primero qué se entiende por «relación».

Ciertamente, una definición formal de «relación» debería comprender relaciones familiares como «$x = y$», «$x < y$», «x es la madre de y» y «x es de más edad que y». Si X es un conjunto, ¿qué es lo que determina una relación entre los pares de elementos de X? Lo que se necesita, evidentemente, es una ley para determinar si, para dos elementos dados cualesquiera x e y de X, x está o no en la correspondencia dada con y. Esa ley R se la llama una **relación** (binaria) en X. Si queremos ser más precisos se puede proceder como sigue. Sea $X \times X$ el conjunto de todos los pares ordenados (x, y) de elementos en X.

Una relación binaria en X es una función R de $X \times X$ en el conjunto $\{0, 1\}$. Esto es, R asigna a cada par ordenado (x, y) el valor 1 o 0. La idea es que si $R(x, y) = 1$, entonces x está en la correspondencia dada con y, y si $R(x, y) = 0$, no lo está.

Si R es una relación binaria en el conjunto X es conveniente escribir xRy cuando $R(x, y) = 1$. Una relación binaria R se dice

(1) **reflexiva**, si xRx para todo x de X;
(2) **simétrica**, si yRx toda vez que xRy;
(3) **transitiva**, si xRz toda vez que xRy y yRz.

Una **relación de equivalencia** en X es una relación binaria en X que es reflexiva, simétrica y transitiva.

Ejemplo 5. (a) En cualquier conjunto la igualdad es una relación de equivalencia. En otras palabras, si xRy quiere decir $x = y$, entonces R es una relación de equivalencia. En efecto, $x = x$, si $x = y$ entonces $y = x$, si $x = y$ e $y = z$ entonces $x = z$. La relación «$x \neq y$» es simétrica, pero no es reflexiva ni transitiva.

(b) Sea X el conjunto de los números reales y supóngase que xRy quiere decir $x < y$. Entonces R no es una relación de equivalencia: es transitiva, pero no es reflexiva ni simétrica. La relación «$x \leq y$» es reflexiva y transitiva, pero no es simétrica.

(c) Sea E el plano euclidiano y sea X el conjunto de todos los triángulos del plano E. Entonces la congruencia es una relación de equivalencia en X; esto es, «$T_1 \cong T_2$» (T_1 es congruente con T_2) es una relación de equivalencia en el conjunto de todos los triángulos en el plano.

(d) Sea X el conjunto de todos los enteros

$$\ldots, -2, -1, 0, 1, 2, \ldots.$$

Sea n un entero positivo fijo. Se define una relación R_n en X por: xR_ny si, y solo si, $(x - y)$ es divisible por n. La relación R_n se llama **congruencia módulo** n. En vez de xR_ny es corriente escribir

$$x \equiv y, \text{ mód } n \qquad (x \text{ es congruente a } y \text{ módulo } n)$$

cuando $(x - y)$ es divisible por n. Para cada entero positivo n, la congruencia módulo n es una relación de equivalencia en el conjunto de los enteros.

(e) Sean X e Y conjuntos y f una función de X en Y. Se define una relación R en X por: x_1Rx_2 si, y solo si, $f(x_1) = f(x_2)$. Es fácil verificar que R es una relación de equivalencia en el conjunto X. Como se verá, este ejemplo particular comprende en realidad todas las relaciones de equivalencia.

Supóngase que R es una relación de equivalencia en el conjunto X. Si x es el elemento de X, $E(x; R)$ representa el conjunto de todos los elementos y de X tales que xRy. Este conjunto $E(x; R)$ se llama la **clase de equivalencia de** x (para la relación de equivalencia R). Como R es una relación de equivalencia, las clases de equivalencia tienen las siguientes propiedades:

(1) Todo $E(x; R)$ es no vacío; en efecto, como xRx, el elemento x pertenece a $E(x; R)$.
(2) Sean x e y elementos de X. Como R es simétrica, y pertenece a $E(x; R)$ si, y solo si, x pertenece a $E(y; R)$.
(3) Si x e y son elementos de X, las clases de equivalencia $E(x; R)$ y $E(y; R)$ o son idénticas o no tienen elementos en común. Primero supóngase que xRy. Sea z cualquier elemento de $E(x; R)$, es decir, un elemento de X tal que xRz. Como R es simétrica, se tiene que zRx. Por hipótesis, xRy, y puesto que R es transitiva, se tiene zRy o yRz. Esto muestra que todo elemento de $E(x; R)$ es elemento de $E(y; R)$. Por la simetría de R, del mismo modo podemos ver que todo elemento de $E(y; R)$ es elemento de $E(x; R)$; luego $E(x; R) = E(y; R)$. Ahora bien, se afirma que si la relación xRy no es válida, entonces $E(x; R) \cap E(y; R)$ es vacío. En efecto, si z está en ambas de estas clases de equivalencia tenemos que xRz e yRz, o sea xRz y zRy, luego xRy.

Si $\mathfrak{F}$ representa una familia de clases de equivalencia para la relación de equivalencia R, vemos que: (1) todo conjunto de la familia $\mathfrak{F}$ es no vacío, (2) todo elemento x de X pertenece a uno, y solo uno, de los conjuntos de la familia $\mathfrak{F}$, (3) xRy si, y solo si, x e y pertenecen al mismo conjunto en la familia $\mathfrak{F}$. En resumen, la relación de equivalencia R subdivide a X en la unión de una familia de subconjuntos (no vacíos) que no se solapan. El razonamiento es también válido a la inversa. Supóngase que $\mathfrak{F}$ es una familia de subconjuntos de X que satisfacen las condiciones (1) y (2) anteriores. Si se define una relación R por (3), entonces R es una relación de equivalencia sobre X y $\mathfrak{F}$ es la familia de clases de equivalencia segun R.

Ejemplo 6. Veamos cuáles son las clases de equivalencia según las relaciones de equivalencia en el Ejemplo 5.

(a) Si R es la igualdad en el conjunto X. entonces la clase de equivalencia del elemento x es simplemente el conjunto $\{x\}$, cuyo único elemento es x.
(b) Si X es el conjunto de todos los triángulos en el plano y R es la relación de congruencia, todo lo que se puede decir en principio es que la clase de equivalencia del triángulo T consta de todos los triángulos que son congruentes con T. Uno de los objetivos de la geometría plana es dar otras descripciones de estas clases de equivalencia.
(c) Si X es el conjunto de los enteros y R_n es la relación «congruencia módulo n», entonces hay precisamente n clases de equivalencia. Cada entero x se puede expresar unívocamente en la forma $x = qn + r$, donde q y r son enteros y $0 \leq r \leq n - 1$. Esto muestra que cada x es congruente módulo n a uno precisamente de los n enteros $0, 1, 2, \ldots, n - 1$. Las clases de equivalencia son

$$\begin{aligned}
E_0 &= \{\ldots, -2n, -n, 0, n, 2n, \ldots\} \\
E_1 &= \{\ldots, 1 - 2n, 1 - n, 1 + n, 1 + 2n, \ldots\} \\
\vdots &= \qquad\qquad \vdots \\
E_{n-1} &= \{\ldots, n - 1 - 2n, n - 1 - n, n - 1, n - 1 + n, \\
&\qquad\qquad n - 1 + 2n, \ldots\}.
\end{aligned}$$

(d) Supóngase que X e Y son conjuntos, f una función de X en Y y R es la relación de equivalencia definida por x_1Rx_2 si, y solo si $f(x_1) = f(x_2)$. Las clases de equivalencia según R son los mayores subconjuntos de X en los que f es «constante». Otra descripción de las clases de equivalencia es ésta: están en correspondencia biunívoca con los elementos de la imagen de f. Si y está en la imagen de f, el conjunto de todos los x de X tales que $f(x) = y$ es una clase de equivalencia según R; y esto define una correspondencia biunívoca entre los elementos de la imagen de f y las clases de equivalencia de R.

Un comentario más respecto a las relaciones de equivalencia. Dada una relación de equivalencia en X, sea $\mathfrak{F}$ la familia de las clases de equivalencia según R. La asociación de la clase de equivalencia $E(x; R)$ con el elemento x define una función f de X en $\mathfrak{F}$ (en realidad, sobre $\mathfrak{F}$):

$$f(x) = E(x; R).$$

Esto muestra que R es la relación de equivalencia asociada a una función cuyo dominio es X, como en el Ejemplo 5(e). Lo que esto indica es que toda relación de equivalencia en el conjunto X está determinada como sigue: Se tiene una ley (función) f que asocia a cada elemento x de X un objeto $f(x)$, y xRy si, y solo si $f(x) = f(y)$. Ahora debe pensarse en $f(x)$ como una propiedad de x, de modo que lo que la relación de equivalencia hace (someramente hablando) es reunir a todos aquellos elementos de X que tienen esta propiedad en común. Si el objeto $f(x)$ es la clase de equivalencia de x, entonces todo lo que se ha dicho es que la propiedad común de los elementos de una clase de equivalencia es que pertenecen a la misma clase. Claro es que esto no dice mucho. Por lo general, hay muchas funciones diferentes, f, que determinan, cómo se hace antes, la relación de equivalencia dada, y un objetivo en el estudio de las relaciones de equivalencia es hallar una f tal que dé una descripción elemental y con sentido de la relación de equivalencia. En la Sección A.5 veremos cómo se hace esto para algunas relaciones de equivalencia especiales que aparecen en el álgebra lineal.

A.4. Espacios cocientes

Sea V un espacio vectorial sobre el cuerpo F y sea W un subespacio de V. Hay en general muchos subespacios W' que son complementarios de W, es decir, subespacios con la propiedad de que $V = W \oplus W'$. Si tenemos un producto interior sobre V, y W es de dimensión finita, hay un subespacio particular que se podría llamar, probablemente, subespacio complementario «natural» de W. Este es el complemento ortogonal de W. Pero si V no tiene otra estructura fuera de su estructura de espacio vectorial, no hay modo de seleccionar un subespacio W' que se pudiera llamar subespacio complementario natural de W. Sin embargo, se puede construir a partir de V y W un espacio vectorial V/W, llamado «cociente» de V y W, que hará de complemento natural de W. Este espacio cociente no es un subespacio de V y, por tanto, no puede ser un subespacio complementario de W; pero es un espacio vectorial definido solo en tér-

minos de V y W y tiene la propiedad de que es isomorfo a cualquier subespacio W' que sea complementario de W.

Sea W un subespacio de un espacio vectorial V. Si α y β son vectores en V, se dice que α es **congruente con β módulo** W si el vector $(\alpha - \beta)$ está en el subespacio W. Si α es congruente con β módulo W, se escribirá

$$\alpha \equiv \beta, \qquad \text{mód } W.$$

Ahora bien, la congruencia módulo W es una relación de equivalencia sobre V.

(1) $\alpha \equiv \alpha$, mód W, pues $\alpha - \alpha = 0$ está en W.

(2) Si $\alpha \equiv \beta$, mód W, entonces $\beta \equiv \alpha$, mód W. En efecto, como W es un subespacio de V, el vector $(\alpha - \beta)$ está en W si, y solo si, $(\beta - \alpha)$ está en W.

(3) Si $\alpha \equiv \beta$, mód W y $\beta \equiv \gamma$, mód W, entonces $\alpha \equiv \gamma$, mód W. En efecto, si $(\alpha - \beta)$ y $(\beta - \gamma)$ están en W, entonces $\alpha - \gamma = (\alpha - \beta) + (\beta - \gamma)$ está en W.

Las clases de equivalencia de esta relación de equivalencia se llaman **clases laterales** de W. ¿Cuál es la clase de equivalencia (clase lateral) de un vector α? Consta de todos los vectores β de V tales que $(\beta - \alpha)$ está en W; esto es, todos los vectores β de la forma $\beta = \alpha + \gamma$, con γ en W. Por esta razón la clase lateral del vector α se representa por

$$\alpha + W.$$

Es apropiado pensar que la clase lateral de α con respecto a W es el conjunto de los vectores que se obtienen por traslación del subespacio W por el vector α. Para ilustrar estas clases laterales, el lector podría pensar en el siguiente caso especial. Sea V el espacio R^2 y sea W un subespacio unidimensional de V. Si se considera V como un plano euclidiano, W es una recta por el origen. Si $\alpha = (x_1, x_2)$ es un vector en V, la clase lateral $\alpha + W$ es la recta que pasa por el punto (x_1, x_2) y es paralela a W.

La colección de todas las clases laterales de W se representará por V/W. Definimos ahora una adición vectorial y una multiplicación por un escalar en V/W como sigue:

$$\begin{aligned}(\alpha + W) + (\beta + W) &= (\alpha + \beta) + W\\ c(\alpha + W) &= (c\alpha) + W.\end{aligned}$$

En otras palabras, la suma de la clase lateral de α y la clase lateral de β es la clase lateral de $(\alpha + \beta)$ y el producto del escalar c y de la clase lateral de α es la clase lateral del vector $c\alpha$. Ahora bien, muchos vectores diferentes de V tendrán la misma clase lateral con respecto a W y, por tanto, se debe verificar que la suma y el producto anterior dependen solo de las clases laterales que intervienen. Lo que esto quiere decir es que debemos demostrar lo siguiente:

(a) Si $\alpha \equiv \alpha'$, mód W y $\beta \equiv \beta'$, mód W, entonces

$$\alpha + \beta \equiv \alpha' + \beta', \qquad \text{mód } W.$$

(2) Si $\alpha \equiv \alpha'$, mód W, entonces $c\alpha \equiv c\alpha'$, mód W.

Estos hechos son fáciles de verificar. (1) Si $\alpha - \alpha'$ está en W y $\beta - \beta'$ está en W, entonces como $(\alpha + \beta) - (\alpha' + \beta') = (\alpha - \alpha') + (\beta - \beta')$, vemos que $\alpha + \beta$ es congruente con $\alpha' + \beta'$ módulo W. (2) Si $\alpha - \alpha'$ está en W y c cualquier escalar, entonces $c\alpha - c\alpha' = c(\alpha - \alpha')$ está en W.

Es ahora fácil de verificar que V/W, con la adición vectorial y la multiplicación por escalares definidos anteriormente, es un espacio vectorial sobre el cuerpo F. Debemos comprobar directamente cada uno de los axiomas para un espacio vectorial. Cada una de las propiedades de la adición vectorial y de la multiplicación por escalares se desprenden de la correspondiente propiedad de las operaciones en V. Habría que decir que el vector nulo de V/W será la clase lateral del vector nulo de V. Es decir, W es el vector nulo de V/W.

El espacio vectorial V/W se llama **cociente** (o diferencia) de V y W. Hay una transformación lineal natural Q de V sobre V/W. Está definida por $Q(\alpha) = \alpha + W$. Debemos notar que se han definido las operaciones en V/W justamente para que esta transformación Q sea lineal. Obsérvese que el espacio nulo de Q es exactamente el subespacio W. Se llama a Q la **transformación cociente** (o aplicación cociente) de V sobre V/W.

La relación entre el espacio cociente V/W y los subespacios de V que son complementarios de W puede ser ahora establecida como sigue.

Teorema. *Sea W un subespacio del espacio vectorial V y sea Q la transformación cociente de V sobre V/W. Supóngase que W' es un subespacio de V. Entonces $V = W \oplus W'$ si, y solo si, la restricción de Q a W' es un isomorfismo de W' sobre V/W.*

Demostración. Supóngase que $V = W \oplus W'$. Esto quiere decir que cada vector α de V está unívocamente expresado en la forma $\alpha = \gamma + \gamma'$, con γ en W y γ' en W'. Entonces $Q\alpha = Q\gamma + Q\gamma' = Q\gamma'$, esto es, $\alpha + W = \gamma' + W$. Esto muestra que Q aplica W' sobre V/W, es decir, que $Q(W') = V/W$. También Q es inyectiva en W'; en efecto, supóngase que γ_1' y γ_2' sean vectores en W' y que $Q\gamma_1' = Q\gamma_2'$. Entonces $Q(\gamma_1' - \gamma_2') = 0$, con lo que $\gamma_1' - \gamma_2'$ están en W. Este vector está también en W', que es disjunto de W; luego $\gamma_1' - \gamma_2' = 0$. La restricción de Q a W' es, por tanto, una transformación lineal inyectiva de W' sobre V/W.

Supóngase que W' es un subespacio de V tal que Q sea inyectiva en W' y que $Q(W) = V/W$. Sea α un vector de V. Entonces existe un vector γ' de W' tal que $Q\gamma' = Q\alpha$; es decir, $\gamma' + W = \alpha + W$. Esto quiere decir que $\alpha = \gamma + \gamma'$ para algún vector γ en W. Por tanto, $V = W + W'$. Para ver que W y W' son disjuntos, supóngase que γ está en W y W'. Como γ está en W, se tiene $Q\gamma = 0$. Pero Q es inyectiva en W', y entonces debemos tener que $\gamma = 0$ Por tanto, tenemos que $V = W \oplus W'$. ∎

Lo que realmente dice este teorema es que W' es complementario de W si, y solo si, W' es un subespacio que contiene exactamente un elemento de cada clase lateral de W. Muestra que cuando $V = W \oplus W'$, la transformación

cociente Q «identifica» W' con V/W. En resumen, $(W \oplus W')/W$ es isomorfo de W' de un modo «natural».

Otro hecho obvio debemos notar. Si W es un subespacio del espacio vectorial de dimensión finita V, entonces

$$\dim W + \dim (V/W) = \dim V.$$

Se puede ver esto del teorema anterior. Tal vez es más fácil observar que lo que esta fórmula dimensional dice es que

$$\text{nulidad } (Q) + \text{rango } (Q) = \dim V.$$

No es nuestro propósito hacer aquí un tratamiento detallado de los espacios cocientes. Pero hay un resultado fundamental que se demostrará.

Teorema. *Sean V y Z espacios vectoriales sobre el cuerpo F. Supóngase que T es una transformación lineal de V sobre Z. Si W es el espacio nulo de T, entonces Z es isomorfo a V/W.*

Demostración. Definimos una transformación U de V/W en Z por $U(\alpha + W) = T\alpha$. Debemos verificar que U está bien definida, es decir, que si $\alpha + W = \beta + W$, entonces $T\alpha = T\beta$. Esto se desprende del hecho de que W es el espacio nulo de T; en efecto, $\alpha + W = \beta + W$ quiere decir que $\alpha - \beta$ está en W, y esto sucede si, y solo si, $T(\alpha - \beta) = 0$. Esto muestra no solo que U está definida, sino que también U es inyectiva.

Es ahora fácil comprobar que U es lineal y aplica V/W sobre Z, pues T es una transformación lineal de V sobre Z. ∎

A.5. Relaciones de equivalencia en Algebra Lineal

Consideraremos algunas de las relaciones de equivalencia que aparecen en el texto de este libro. Este es apenas una muestra de tales relaciones.

(1) Sean m y n enteros positivos y F un cuerpo. Sea X el conjunto de todas las matrices $m \times n$ sobre F. Entonces la equivalencia por filas es una relación de equivalencia en el conjunto X. La afirmación «A es equivalente por filas a B» quiere decir que A se puede obtener de B por una sucesión finita de operaciones elementales por filas. Si escribimos $A \sim B$ por «A es equivalente por fila a B», entonces no es difícil comprobar las propiedades (i) $A \sim A$; (ii) si $A \sim B$, entonces $B \sim A$; (iii) si $A \sim B$ y $B \sim C$, entonces $A \sim C$. ¿Qué se sabe de esta relación de equivalencia? En realidad se sabe mucho. Por ejemplo, se sabe que $A \sim B$ si, y solo si, $A = PB$ para cierta matriz inversible $m \times m$, P; o $A \sim B$ si, y solo si, el sistema homogéneo de ecuaciones lineales $AX = 0$ y $BX = 0$ tienen las mismas soluciones. También se tiene información explícita respecto a las clases de equivalencia para esta relación. Cada matriz $m \times n$, A, es equivalente por filas a una, y solamente una, matriz escalón reducida por filas. Lo que esto dice es que cada clase de equivalencia, según esta relación,

contiene precisamente una matriz escalón reducida por filas, R; la clase de equivalencia determinada por R consta de todas las matrices $A = PR$ donde P es una matriz inversible $m \times m$. También se puede pensar en esta descripción de las clases de equivalencia del siguiente modo. Dada una matriz $m \times n$, A, tenemos una ley (función) f que asocia con A la matriz escalón reducida por filas $f(A)$ que es equivalente por filas a A. La equivalencia por filas está completamente determinada por f. En efecto, $A \sim B$ si, y solo si, $f(A) = f(B)$, es decir, si, y solo si, A y B tienen la misma forma escalón reducida por filas.

(2) Sea n un entero positivo y F un cuerpo. Sea X el conjunto de todas las matrices $n \times n$ sobre F. Entonces la semejanza es una relación de equivalencia en X; toda matriz $n \times n$, A, es semejante a sí misma; si A es semejante a B, entonces B es semejante a A; si A es semejante a B y B es semejante a C, entonces A es semejante a C. Sabemos también bastante de esta relación de equivalencia. Por ejemplo, A es semejante a B si, y solo si, A y B representan el mismo operador lineal sobre F^n en (posiblemente) diferentes bases ordenadas. Pero se sabe algo más profundo que esto. Toda matriz $n \times n$, A, sobre F es semejante (sobre F) a una, y solo a una, matriz que está en la forma racional (Capítulo 7). En otras palabras, cada clase de equivalencia, según la relación de semejanza, contiene precisamente una matriz que está en la forma racional. Una matriz en la forma racional está determinada por un k-tuple $(p_1, \ldots, p_n)$ de polinomios mónicos que tienen la propiedad de que p_{j+1} divide a p_j, $j = 1, \ldots, k-1$. Así, se tiene una función f que asocia a cada matriz $n \times n$, A, un k-tuple $f(A) = (p_1, \ldots, p_k)$ que satisface la condición de divisibilidad de que p_{j+1} divide a p_j. Y, A y B son semejantes si, y solo si, $f(A) = f(B)$.

(3) He aquí un caso especial del anterior Ejemplo 2. Sea X el conjunto de las matrices 3×3 sobre el cuerpo F. Se considera la relación de semejanza en X. Si A y B son matrices 3×3 sobre F, entonces A y B son semejantes si, y solo si, tienen el mismo polinomio característico y el mismo polinomio minimal. Asociado a cada matriz 3×3, A, se tiene un par de polinomios mónicos (p, q) que satisfacen

(a) $\text{grd}\, f = 3$;
(b) p divide f;

siendo f el polinomio característico para A y p el polinomio minimal para A. Dados los polinomios mónicos f y p sobre F, que cumplen (a) y (b), es fácil encontrar una matriz 3×3 sobre F que tiene a f y a p como sus polinomios característico y minimal, respectivamente. Lo que todo esto significa es que: Si se considera la relación de semejanza en el conjunto de las matrices 3×3 sobre F, las clases de equivalencia están en correspondencia biunívoca con los pares ordenados (f, p) de polinomios mónicos sobre F que satisfacen (a) y (b).

A.6. El axioma de elección

Someramente hablando, el axioma de elección es una regla (o principio) del pensamiento que dice que, dada una familia de conjuntos no vacíos, se

puede elegir un elemento de cada conjunto. Para ser más precisos, supóngase que se tenga un conjunto de índices A y que para cada α de A se tenga un conjunto asociado S_α no vacío. El «elegir» o «seleccionar» un elemento de cada S_α quiere decir dar una regla f que asocie a cada α un elemento $f(\alpha)$ en el conjunto S_α. El axioma de elección dice que esto es posible; es decir, dada la familia de conjuntos $\{S_\alpha\}$, existe una función f de A en

$$\bigcup_\alpha S_\alpha$$

tal que $f(\alpha)$ está en S_α para todo α. Este principio es aceptado por la mayoría de los matemáticos, pese a que surgen muchas situaciones en las que está lejos de ser claro cómo se pueda hallar una función explícita f.

El axioma de elección tiene consecuencias sorprendentes. La mayoría de las cuales tienen poca o ninguna influencia en el material de este libro; sin embargo, es digna de mencionar una: todo espacio vectorial tiene una base. Por ejemplo, el cuerpo de los números reales tiene una base, considerado como espacio vectorial sobre el cuerpo de los racionales. Es decir, existe un subconjunto S de R que es linealmente independiente sobre el cuerpo de los racionales y tiene la propiedad de que cada número real es una combinación lineal racional de un número finito de elementos de S. No nos detendremos aquí a derivar este espacio vectorial que resulta de la aplicación del axioma de elección. Para una demostración, el lector puede consultar el libro de Kelley citado en la bibliografía.

Bibliografía

Halmos, P., *Finite-Dimensional Vector Spaces*, D. Van Nostrand Co., Princeton, 1958.

Jacobson, N., *Lectures in Abstract Algebra*, II, D. Van Nostrand Co., Princeton, 1953.

Kelley, John L., *General Topology*, D. Van Nostrand Co., Princeton, 1955.

MacLane, S. and Birkhoff, G., *Algebra*, The Macmillan Co., New York, 1967.

Schreier, O. and Sperner, E., *Introduction to Modern Algebra and Matrix Theory*, 2nd Ed., Chelsea Publishing Co., New York, 1955.

Van der Waerden, B. L., *Modern Algebra* (dos volúmenes), Rev. Ed., Frederick Ungar Publishing Co., New York, 1969.

Indice